UNIVERSIDAD DE MURCIA

FACULTAD DE LETRAS

Foreign Language Pronunciation training with affordable and easily accessible Technologies: Podcasts, Smartphone Apps and Social Networking Services (Twitter)

Enseñanza de la Pronunciación de la Lengua Extranjera a través de Tecnologías asequibles y de fácil acceso: Podcasts, Aplicaciones Móviles y Redes Sociales (Twitter)

D. Jonás Fouz González

2015

UNIVERSIDAD DE MURCIA

FACULTAD DE LETRAS

Foreign language pronunciation training with affordable and easily accessible technologies: podcasts, smartphone apps and social networking services (Twitter)

Enseñanza de la pronunciación de la lengua extranjera a través de tecnologías asequibles y de fácil acceso: podcasts, aplicaciones móviles y redes sociales (Twitter)

Directores:

Dr. Rafael Monroy Casas

Dr. José Antonio Mompeán González

D. Jonás Fouz González

2015

To my family

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgements

I owe my interest in pronunciation to the Lecturers and Professors of *Prácticas de pronunciación, Fonética inglesa* and *Entonación inglesa* in the degree in English Philology at the University of Murcia. It is for this reason that I am profoundly grateful and honoured to have had two such incredible professionals – José Antonio Mompeán González and Rafael Monroy Casas – supervising my PhD dissertation throughout these years. I heartily thank you both for your constant support and help whenever I required it, for your extremely valuable suggestions in the design of the studies and for your meticulous feedback in earlier versions of this dissertation. Thanks, Rafael, for those 'private lessons' helping me grasp the essence of what research is, and for your words of encouragement at times of uncertainty. Thanks, Jose, for your infinite patience and guidance over the years, for your availability and for your friendship. I have learnt from you more than you can imagine.

I am most grateful to the people who have made this dissertation bearable at times that were exceedingly difficult. Thanks, Ángela, Lali and Eugenio for your invaluable help labelling, editing and organising the myriad of audios that haunted me for months. Speaking of audios, I do not have words to thank Pilar Mompeán Guillamón and Ana Rosa Sánchez Muñoz for their readiness to help, their availability and their endless patience with those 'two days' that turned out to be more than two months. I would also like to thank Aurora González Vidal and Antonio Maurandi López for their assistance and guidance with the statistical analyses in the three empirical studies. Thank you all for your time and willingness to help, without you, I simply would have not been able to make it.

My thanks also go to Kate Chedgzoy, Scott Windeatt and Cristina Dye for making my research stay in Newcastle possible. I am also deeply grateful to Santi, Christian and Ralf for making me feel at home and for their support all these years, and to Ian for acting as a voice model for the app.

I am also indebted to Hugo, Alejandro and Jacobo for helping me with the app over the summer, and to Valerie Biggs, Rachel Green, Paul Clarkson, Sean Lewis and Catherine Staveley for their help with the stimuli for the perception tests. Thanks to Mar Vilar, Eduardo Saldaña, Raquel Morcillo and Pilar Martínez for helping me recruiting participants for the studies in the dissertation, and to Lourdes Cerezo for believing in me and encouraging me to pursue this career. Additionally, I would also like to thank the participants who volunteered to take part in these studies.

The last year of this dissertation has been particularly challenging for various reasons, but my colleagues at the Catholic University in Murcia have made the countless hours of work feel less of a burden and they have made work feel like home. Thank you all for your support throughout these tough months, but especially Diana, Pilar, Ana, Imelda, Laura, Gloria, Jorge, Raquel and José Andrés for your constant enquiries about my progress. I would also like to thank Alba, Jorge and Sara for their help with some challenging translations, and special thanks are due to Imelda and Brendan for their patience and assistance with my many questions about language. Last, but not least, I am also deeply grateful to Thomas Schmidt for his constant support, especially during these last stressful months.

Finally, I would like to give a very special thanks to my family from Murcia, Zamora and Lugo, for being always there for me, and to Lali, for her patience and wholehearted support over the years and her willingness to help in everything I needed no matter how busy she was with her own dissertation.

Resumen

El aprendizaje de una lengua extranjera (LE) se ha convertido en algo a lo que la mayoría de la gente se enfrenta a lo largo de su vida, ya sea por interés personal, perspectivas de mejorar un currículum, o como parte de la etapa de educación obligatoria. Una de las razones principales por las que la gente estudia un idioma es para poder comunicarse con los hablantes de esa lengua; y en el caso del inglés, garantizando la posibilidad de interactuar con hablantes de cualquier parte del mundo, dado su papel como *lingua franca* en la sociedad actual.

A pesar de que el aprendizaje de una LE conlleva el dominio de diversos aspectos, como la gramática, el léxico o la morfosintaxis, una de las competencias más importantes que los aprendices deben adquirir es la pronunciación, especialmente para la comunicación oral. Esto se debe al hecho de que un hablante con perfecto dominio de la gramática y del léxico puede ser completamente ininteligible si su pronunciación no es adecuada (Brown, 1991; Hinofotis y Bailey, 1980).

Desafortunadamente para los estudiantes de idiomas la pronunciación es una de las competencias más difíciles de adquirir, ya que no depende únicamente del nivel de esfuerzo dedicado ni de las capacidades cognitivas de los alumnos, sino también de habilidades perceptuales y psicomotoras (Pennington, 1998). Un ejemplo de la dificultad que supone adquirir esta competencia es que, a pesar de la existencia de numerosos casos en que los alumnos que alcanzan un nivel nativo o casi nativo en otros aspectos del idioma, como el léxico o la gramática, son muy pocos los que logran alcanzar un nivel nativo de pronunciación (Moyer, 1999).

La sabiduría popular ha mantenido durante muchos años la postura de que la pronunciación 'no es importante', que los alumnos 'la adquirirán por sí mismos', o que es algo que 'no se puede enseñar', aunque los profesores quieran (Morley, 1994, p. 66). Algunos autores han llegado a recomendar que se entrenase a hablantes nativos a entender el acento extranjero antes que invertir dinero en mejorar la pronunciación de los profesores no nativos en América (Rubin, 1992). Esto se debe, en gran parte, al escaso número de hablantes no nativos que consiguen dominar un idioma hasta el punto de no mostrar rasgo alguno de acento extranjero (Scovel, 1988).

El primer capítulo de esta tesis doctoral hace una revisión de la literatura sobre la enseñanza de lenguas asistida por ordenador, ofreciendo diversos ejemplos que atestiguan

que enseñar pronunciación es posible, incluso después de periodos de entrenamiento que no van más allá de unas horas (véase sección 1.3). Paradójicamente, a pesar del papel fundamental de esta competencia para la comunicación, la pronunciación tiende a relegarse a segundo plano en el aula de LE (Celce-Murcia, Brinton, Goodwin, y Griner, 2010; Setter y Jenkins, 2005), a menudo con el objeto de dedicar más tiempo a competencias como la lectura o la escritura, más dadas a fomentar el éxito académico en culturas que se guían por exámenes centrados en esas destrezas (Setter, 2008). Algunas de las razones por las que se ignora la pronunciación en las aulas de idiomas son las limitaciones de tiempo en el aula (Levis, 2007), la falta de preparación de profesores (Henderson et al., 2015) o la escasez de materiales adecuados (Busà, 2008; Calvo-Benzies, 2013).

Dados los numerosos obstáculos que los hablantes tienen que sortear para mejorar su pronunciación, como la lengua materna, la edad, la timidez, las aptitudes individuales o simplemente la falta de tiempo y materiales (véase sección 1.1.2), muchos autores han apostado por la tecnología para facilitar a los alumnos la adquisición/aprendizaje de esta competencia. A lo largo de la historia, la motivación para desarrollar nuevas herramientas tecnológicas ha surgido por el deseo de ampliar nuestras limitaciones como seres humanos, especialmente las limitaciones físicas y mentales (Levy y Stockwell, 2006). Como señalan Levy y Stockwell, herramientas como el bolígrafo y el papel sirven de apoyo a la memoria a la hora de escribir; el telescopio o el microscopio aumentan nuestro rango de visión; el martillo, nuestra fuerza. En lo que a la pronunciación respecta, la tecnología también presenta un amplio abanico de ventajas a la hora de facilitar su adquisición, especialmente dadas las dificultades perceptuales y motoras de los alumnos ocasionadas principalmente por la interferencia de la lengua materna (Lado, 1957).

Hay dos aspectos fundamentales que los aprendices de una LE deben dominar con respecto a la pronunciación: percepción y producción. Por un lado, los alumnos deben ser capaces de percibir los patrones y sonidos de la lengua extrajera, y por otro, deben ser capaces de realizar los movimientos articulatorios necesarios para obtener una producción adecuada de los mismos. Esto resulta extremadamente difícil para los estudiantes de idiomas, como ya decíamos antes debido a la gran influencia que ejerce su lengua materna, ya que los alumnos tienden a percibir los sonidos de la LE a través de las categorías perceptuales que tienen establecidas para su primera lengua (Kuhl, 1991; Kuhl & Iverson, 1995; Flege, 1987; Flege, 1995).

En este sentido, la tecnología ofrece multitud de posibilidades para facilitar una percepción y producción adecuadas. En cuanto a la percepción, la sección 1.3 del capítulo I hace una revisión de las diferentes posibilidades que ofrece la tecnología para ayudar a los alumnos a percibir y producir la pronunciación de la LE de modo satisfactorio. Los investigadores han explorado el potencial de herramientas de análisis del habla, como por ejemplo los espectrogramas, los visualizadores de contornos entonativos o los oscilogramas para ayudar a los alumnos de LE a mejorar su pronunciación. No obstante, a pesar de la utilidad de estas herramientas para trabajar ciertos aspectos de la pronunciación, éstas han recibido duras críticas por parte de diversos investigadores debido a la dificultad de interpretación del feedback ofrecido por las mismas para el aprendiz de idiomas de a pie, ya que son herramientas concebidas originalmente para el análisis acústico (Levis, 2007; Llisterri, 2001; O'Brien, 2006).

En cuanto a la producción, una tecnología que ha despertado el interés de gran número de investigadores es el reconocimiento de habla automático, puesto que, con las herramientas adecuadas, sería posible evaluar las producciones de los alumnos y ofrecer feedback instantáneo. Este tipo de *feedback* se ha utilizado en numerosas aplicaciones comerciales y programas de aprendizaje a distancia, como *Tell Me More* o *Talk To Me*. Sin embargo, la tecnología actual no es capaz de reconocer correctamente todos los errores que cometen los hablantes no nativos, siendo una de las mayores limitaciones el hecho de que el *feedback* suele ser erróneo (Neri, Cucchiarini, Strik, y Boves, 2002), diciendo a los alumnos que algo está bien pronunciado cuando no es así, o que algo está mal pronunciado cuando en realidad la pronunciación era adecuada (Eskenazi, 2009).

Dado que no existe una herramienta perfecta, capaz de detectar los errores de los alumnos e informarles sobre cómo mejorar, el método adoptado en esta tesis doctoral consiste en entrenar a los alumnos a percibir la pronunciación de la LE de manera adecuada con la intención de ayudarlos a crear modelos perceptuales correctos que les sirvan para autoevaluar su pronunciación. Se cree que, si los alumnos son capaces de percibir la pronunciación de manera apropiada, estarán en una mejor posición para poder evaluar de manera autónoma su producción (Avery y Ehrlich, 1992). En este sentido, la sección 1.1.5 presenta varios estudios que apoyan la importancia de la atención a la forma y la instrucción explícita para la pronunciación (Saito y Lyster, 2012; Saito, 2013). Por eso, el método empleado en los tres estudios de esta tesis consiste en dirigir la atención de los alumnos de manera explícita hacia los elementos formales de la pronunciación, con intención de facilitar que los alumnos perciban y sean conscientes de las diferencias entre la pronunciación de la LE y su L1, intentando fomentar cambios que les lleven a conseguir una producción adecuada en la LE.

Como respuesta a la necesidad de materiales y técnicas que apoyen la enseñanza de la pronunciación, y dadas las limitaciones de ciertas herramientas mencionadas anteriormente, el principal objetivo de esta tesis doctoral es investigar las posibilidades que ofrecen diversas herramientas asequibles y de fácil acceso para optimizar la percepción y producción de una selección de aspectos que se consideran problemáticos para los hablantes de LE. Asimismo, se investigan las opiniones de los alumnos con respecto a estas herramientas así como el grado de utilidad percibido. El capítulo II ofrece una perspectiva general de los estudios de este trabajo. En ese capítulo se expone la justificación de la elección de cada tecnología a evaluar, además de los objetivos de pronunciación fijados para cada estudio. La tesis presenta tres estudios, cada uno centrado en una herramienta distinta. Concretamente, el capítulo IV se centra en podcasts, el capítulo V en aplicaciones móviles y el capítulo VI en redes sociales (Twitter).

Estos tres estudios siguen un diseño de métodos mixtos, utilizando pre-tests y post-tests para evaluar la percepción y la producción de los alumnos en los aspectos meta antes y después de recibir instrucción. Del mismo modo, en cada estudio se enviaron cuestionarios antes de iniciar y al finalizar la instrucción con el objeto de obtener información demográfica sobre los participantes, sobre el grado de familiaridad que tenían con las tecnologías que iban a utilizar, así como sus opiniones sobre el método empleado y la utilidad de cada herramienta para la enseñanza de pronunciación. El capítulo III presenta una explicación detallada de las distintas tareas utilizadas para medir la pronunciación de los participantes. En los estudios 1 y 2 se emplearon tareas de discriminación e identificación para medir la percepción de los alumnos (Beddor y Gottfried, 1995), además de tres tareas de producción para evaluar la producción de los alumnos, tanto en imitación, como en producción controlada y espontánea (véase Morley, 1991, 1994). En el estudio 3 se utilizó una tarea de producción en la cual los alumnos tenían que asociar palabras a colores según sus percepciones sinestésicas, con el objetivo de desviar la atención de los alumnos de la pronunciación de las palabras meta.

Con el fin de evaluar el potencial de las herramientas mencionadas anteriormente de manera empírica, se realizó una selección de aspectos potencialmente problemáticos para los estudiantes de inglés como LE. El estudio 1 se centra en el potencial de los podcasts para mejorar el contraste /s – z/ en inglés y la pronunciación de /b d g/ como oclusivas en posición intervocálica. El estudio 2 explora el potencial de la aplicación *English File Pronunciation* (Oxford University Press, 2012) para mejorar las vocales inglesas /æ/, /ɑː/, /ʌ/ y /ə/, y para el contraste /s – z/. Por último, el estudio 3 analiza el potencial de

Twitter para mejorar la pronunciación de elementos léxicos que se suelen pronunciar de manera inadecuada debido a sus correspondencias entre sonido y grafía o a la presencia de cognados en inglés y español con patrones acentuales diferentes.

Diversos autores han señalado que la investigación de enseñanza de pronunciación asistida por ordenador se ha realizado habitualmente en entornos muy controlados, similares a los de un laboratorio, algo que limita la validez ecológica de los estudios realizados (Lord, 2010; Olson, 2014; Wang y Munro, 2004). Por este motivo, en los estudios de esta tesis doctoral se hizo un esfuerzo por llevar a cabo una instrucción que simulase el modo en que este tipo de herramienta se implementaría en un aula de lengua extranjera real. Excepto para los pre- y post-tests de cada estudio, todo se realizó a distancia en los estudios 2 y 3. No obstante, debido a los resultados obtenidos en el estudio piloto para los podcasts y dada la imposibilidad de controlar la realización de ciertas tareas en el estudio 1 (capítulo IV), algunas de las actividades en ese estudio se completaron de manera presencial.

En el estudio 1 (capítulo IV), se dividió a los participantes en dos grupos que actuaron como grupo control y experimental simultáneamente, ya que cada grupo recibió instrucción sobre un aspecto diferente. El grupo 1 recibió instrucción en el contraste /s – z/ y el grupo 2 en /b d g/. La instrucción se llevó a cabo durante tres semanas. Después de una sesión informativa con instrucción explícita sobre los sonidos meta, los alumnos tenían que realizar tres pasos semanales: prestar atención a los aspectos meta en un podcast, grabar su propio podcast y evaluar los podcasts de sus compañeros. Los resultados muestran que la instrucción tuvo un impacto positivo en la percepción y la producción de los sonidos meta para los que fue entrenado cada grupo. No obstante, aunque hubo diferencias estadísticamente significativas para algunos de los sonidos meta, las diferencias entre grupos no llegaron a ser estadísticamente significativas para todos los sonidos.

En el estudio 2 (capítulo V), se dividió a los participantes en tres grupos. Dos grupos pertenecían a un Grado en Estudios Ingleses (al igual que en el estudio 1) y el tercer grupo estaba constituido por participantes de distintas carreras con asignaturas de inglés como lengua extranjera. Los grupos 1 y 3 actuaron como grupos experimentales y el grupo 2 como grupo control. Sin embargo, una vez finalizada la instrucción para los primeros dos grupos y habiendo completado el post-test, el grupo 2 comenzó a recibir instrucción, actuando por tanto también como grupo experimental (en este caso, sólo para percepción). La instrucción en este estudio consistía en utilizar la aplicación mencionada anteriormente durante dos semanas, aproximadamente unos 20 minutos al día. Los resultados revelan que el entrenamiento con la aplicación fue beneficioso para la percepción y producción de los sonidos meta, aunque al igual que en el estudio 1, las mejoras obtenidas por los grupos experimentales sólo alcanzaron la significación estadística para algunos sonidos.

Por último, el estudio 3 (capítulo VI) consistía en enviar a los participantes un *tweet* diario con explicaciones explícitas y concisas sobre los aspectos mencionados arriba. Los participantes pertenecían al primer curso de un Grado en Medicina. Los participantes del grupo experimental recibían *tweets* sobre los aspectos meta, mientras que los participantes en el grupo control recibían un consejo diario sobre aspectos del inglés no relacionados con la pronunciación (falsos amigos, gramática, et cétera). Los resultados muestran que los participantes que recibieron instrucción a través de Twitter mejoraron su pronunciación de las palabras meta de manera significativa y que mantuvieron estas mejoras incluso un mes después del entrenamiento.

Los estudios descritos anteriormente presentan una serie de limitaciones que es necesario mencionar (véase también el capítulo VII). Una de las limitaciones principales de estos estudios reside en la duración de los mismos, ya que a pesar de que se llevaron a cabo durante varias semanas, la duración del entrenamiento en sí no es superior a tres horas y media en ninguno de ellos. Además, los participantes fueron recompensados académicamente por su participación y eran conscientes de que estaban formando parte de un estudio, lo que limita la fiabilidad de la evaluación, ya que los alumnos consideraban la instrucción algo extra y se corría el riesgo de que participasen en el estudio únicamente para conseguir la recompensa. Finalmente, aunque algunos de los problemas del estudio 3 estaban relacionados con el acento léxico, los estudios en esta tesis doctoral se centran principalmente en los elementos segmentales. Los aspectos meta se eligieron debido al grado de dificultad que suponen para muchos aprendices de LE, dado que suelen estar fosilizados en la interlengua de estudiantes con un nivel bastante avanzado (Monroy, 2001). Se pretendía, por tanto, probar de manera empírica si estas tres herramientas podían promover mejoras en la pronunciación de estos aspectos. No obstante, numerosos autores abogan por la priorización de los elementos suprasegmentales debido a su importancia para la inteligibilidad (Derwing, Munro y Wiebe, 1998; Gilbert, 1994; Pennington, 1989).

Para finalizar, esta tesis doctoral también ofrece algunas direcciones para investigaciones futuras (véanse las secciones correspondientes en los capítulos IV, V, VI y VII). A modo de ejemplo, los estudios 2 y 3 exploran el potencial de las aplicaciones móviles y las redes sociales, pero se centran en una aplicación concreta y en una red social concreta. Por tanto, cabe sugerir que futuras investigaciones analicen el potencial de otras aplicaciones móviles así como de otras redes sociales. Del mismo modo, y en línea con la última limitación mencionada, sería interesante realizar estudios similares explorando el potencial de estas tecnologías para enseñar aspectos suprasegmentales. En este sentido, los podcasts parecen una herramienta ideal para practicar aspectos como el ritmo, la reducción vocálica o la entonación.

Chapter I: Foreign language pronunciation training

Pronunciation is one of the most important competences people need to master when learning a foreign language given its importance for adequate communication. It is by means of the acoustic signal that we can understand and transmit oral messages, and there is a threshold level of pronunciation for non-native speakers below which faulty pronunciations can severely impair communication (Hinofotis & Bailey, 1980). Unfortunately, the acquisition of foreign language (FL) pronunciation does not depend on the learners' effort or declarative knowledge, as it is constrained by various cognitive, perceptual, psychomotor and affective factors (Pennington, 1998). While people attain almost native-like mastery in other aspects of a foreign language, such as grammar or lexis, very few learners attain native-like mastery in pronunciation, even after years of exposure to the foreign language (Moyer, 1999). Furthermore, pronunciation is very often neglected in many FL classrooms due to a number of reasons, such as lack of teacher training (Henderson et al., 2015), scarcity of adequate materials (Busà, 2008), or time constraints that lead teachers to relegate this competence to the background in favour of other skills (Levis, 2007).

An avenue for solution is technology, given the possibilities it offers in order to enhance presentation styles and make materials more accessible to learners – not only in terms of physical accessibility, but also cognitively and psychologically (Pennington, 1996). Technology presents numerous advantages for language learning, such as the possibility to receive immediate feedback, individualised instruction, or unlimited practice (see sections 1.2 and 1.3). In the case of pronunciation, technology proffers a wide range of possibilities in order to enhance the learners' perception and production of FL pronunciation (section 1.3) – something especially advantageous given the perceptual constraints FL learners face as a result of their mother tongue interference (section 1.1.2).

Research investigating computer-assisted pronunciation training has often addressed the possibilities offered by tools originally conceived for phonetic analysis, but these have often been criticised since they are not easily interpretable or accessible for the average language learner (section 1.3). Moreover, research has often been conducted under very controlled, laboratory-like settings which do not resemble the way in which learners would receive instruction in a real situation (Lord, 2010; Wang & Munro, 2004). It is for these reasons that this dissertation investigates the potential of affordable and easily accessible technologies to facilitate FL pronunciation training. More specifically, it focuses on podcasts, smartphone apps, and social networking services (Twitter). The technologies explored here are either free or relatively inexpensive, and they can be easily accessed from a wide range of devices. Furthemore, the studies in this dissertation encourage the use of these technologies off-campus in an attempt to emulate the way learners would use these tools for autonomous practice.

1.1 Introduction: Stressing pronunciation

Learning a foreign language has become something most people do at some point in life for various reasons, be it voluntarily, due to personal interests, with a view to a job or CV requirement, or as part of compulsory education. One of the main reasons to learn a foreign language is being able to communicate with speakers of that language, and in the case of English, guaranteeing the possibility of interacting with the other half of the world who speak it. English has undeniably established itself worldwide as a *lingua franca*, and being intelligible in it has become a goal for many (Setter, 2008).

With the arrival of communicative language teaching (CLT) and the emphasis on communication as the end goal of foreign language instruction, the development of oral skills and the ability to engage in meaningful conversational interaction have been considered the primary goals of foreign language teaching (FLT) in the last decades (Ehsani & Knodt, 1998). Although mastering a language entails a certain control of a number of aspects (grammar, lexis, syntax, etc.), one of the most important competences to be acquired is pronunciation, especially for oral communication. A speaker may be very fluent and have a perfect command of grammar and vocabulary and yet be completely unintelligible due to poor pronunciation (Brown, 1991). Unfortunately for FL learners, given the perceptual, cognitive and psychomotor abilities involved (Pennington, 1998), pronunciation is one of the most difficult competences to master, dooming most adult FL learners to speak with a foreign accent (Scovel, 1988).

Paradoxically enough, despite the vital role it plays in communication, pronunciation has been very often neglected in FLT in favour of other competences (Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Gilbert, 1994; 2010; Henderson et al., 2015; Levis, 2007; Morley, 1994; Setter & Jenkins, 2005). Throughout history, the emphasis received by pronunciation has varied depending on the linguistic theory in vogue at the time; with some approaches completely ignoring it (e.g. Grammar Translation) and some contemplating it as one of the essentials of adequate language learning (e.g. Audio-lingual method, Direct Method, the Silent Way).[1] Since the arrival of Communicative Language Teaching, there seems to be some consensus that communication is the ultimate goal of language learning, at least for the majority of FL learners. Research suggests that there is a threshold level of pronunciation for non-native speakers below which faulty pronunciations can severely impair the communication process (Hinofotis & Bailey, 1980). Below this level, learners will not be understood regardless of their grammatical and lexical mastery of the language (Celce-Murcia et al., 2010). Hence, this communicative view of language should, in theory, bring pronunciation to a privileged position given the pivotal role of pronunciation in oral communication. However, it was precisely this emphasis on meaningful communicative contexts that marked yet another decline in pronunciation instruction, given that it was still associated to behaviouristic drilling of sound contrasts and as something related to linguistic accuracy, rather than fluency and communication (Celce-Murcia et al., 2010; Isaacs, 2009).

Nowadays, the role of pronunciation as an essential element for oral communication is not in question any longer. It is by means of the acoustic signal that we can transmit and decode oral messages; if pronunciation is not intelligible, communication is impossible. Moreover, poor pronunciation does not only hinder communication. A number of unfavourable unconscious attitudes towards foreign speakers may be triggered simply by the way people speak. As Beebe (1978, p. 121) points out, 'pronunciation – like grammar, syntax, and discourse organization – *communicates*. […] The very act of pronouncing, not just the words we transmit, [is] an essential part of what we communicate about ourselves as people'.[2] Personal qualities may be negatively judged due to anticipatory apprehensive reactions on the part of listeners, thinking that they will not be able to understand the interaction as it moves along (Morley, 1994). Recurrent mispronunciations may hinder comprehensibility (Derwing, Munro, & Carbonaro, 2000; Gynan, 1985), impede the correct

---

[1] See Larsen-Freeman (2000) or Sánchez (2009) for a comprehensive account of the different methods and how they dealt with pronunciation.

[2] Cited from Morley (1991, p. 10). In her study, Beebe reported that native speakers often label non-native speaker pronunciation errors derisively, as sounding comical, cute, incompetent, not serious, childish, et cetera.

identification of words if lexical stress is misplaced (Cutler, 1984; Field, 2005; Hahn, 2004), give the impression that a speaker is bored or apathetic if there is no variation in intonation (Kenworthy, 1987; Pickering, 2001), or even sound ungrammatical as a result of different types of segmental and suprasegmental errors (Kennedy, 2015). Furthermore, listeners may become distracted or irritated by recurrent mispronunciations and hesitation (Fayer & Krasinski, 1987). Lack of intelligibility due to a poor pronunciation may occasionally even cause listeners to 'switch off' and, in more extreme cases, even avoid further interactions with the speaker (Singleton, 1995), perhaps due to the anxiety presented by having to face a conversation with someone you cannot understand. This does not necessarily stem from any kind of prejudice, but perhaps due to the listeners' lack of confidence in their abilities to interact with FL learners as a consequence of their little experience interacting with FL learners, as suggested by Derwing and Munro (2009).

Depending on the situation, a poor pronunciation may or may not have serious consequences for the speaker. With regard to intonation, for instance, using the wrong intonational cues may lead to confusion when speakers do not rely on the same cues (Chun, 1988). Chun mentions the use of intonation as an indicator of turn-taking as one of the most likely causes of misunderstanding. Intonation is one of the indicators that mark whether a speaker wants to continue a turn in the conversation or is willing to give up the turn. Thus, if a speaker uses a rising intonation because (s)he expects a reply but this reply does not materialise, the listener might be considered rude because (s)he remained silent when (s)he was supposed to speak. This could be interpreted as the listener deliberately not answering a question, for example. Similarly, using the wrong sentence stress may also sound rude or lead to misinterpretations. Kelly (2000, p. 11) illustrates how shifting the stress from one word to another may give the impression of impatience. She describes how the request *Do you mind if I open the window?* was pronounced by a student with the most prominent stress on *open*, while it should have been on *window*, since it is the last lexical item in the intonation unit. By stressing *open* instead of *window*, the utterance could have been interpreted as being a second request, and therefore make the speaker sound impatient and rude (possibly because the first request may have not been heard by the listener). In a conversation with a friend, the anecdote would probably not transcend, but at a job interview, this may have life-changing consequences.

As Setter and Jenkins (2005, p. 1) write, 'at the affective level, it is through the way we speak, and above all, by means of our accent, that we project our regional, social and ethnic identities'. The way we speak conveys a great deal of information about ourselves (e.g. attitudes, mood, personality). However, it is important to draw a distinction between inadequate pronunciation and accent. Inadequate pronunciation can hinder or even make comprehension impossible (Hinofotis & Bailey, 1980), whereas a foreign accent simply implies divergences from the native norm along a range of segmental and suprasegmental dimensions (Flege, 1995). Early pronunciation teaching approaches placed a very high value on accuracy and native-like proficiency, and consequently, any traits of foreign accent were considered undesirable. Unfortunately, even though research has shown that foreign accent cannot be equated to lack of intelligibility (Munro & Derwing, 1995), numerous studies show that speakers are commonly undermined as a result of their foreign accent.

Speakers showing traits of foreing-accented speech have been judged as having a lower status (Brennan & Brennan, 1981), lacking credibility (Lev-Ari & Keysar, 2010), or even caused irritability (Fayer & Krasinski, 1987). Furthermore, foreign accents have been reported to be detrimental to aspects that have serious consequences in life, such as employers who are resistant to foreign accented speech (Sato, 1991), or even cases in which up to 42% of students have dropped out from courses when they found out they were taught by non-native teaching assistants (Rubin & Smith, 1990).[3]

In spite of the above, the potential negative consequences of foreign accentedness can be palliated if speech is intelligible (see e.g. Bresnahan, Ohashi, Nebashi, Liu, & Shearman, 2002; Rubin, 1992). Moreover, as noted by Cortés Moreno (2000), foreign accented speech may bring about some advantages too, like warning interlocutors that they are talking to a foreigner (therefore encouraging them to be more tolerant), or making them adapt their oral production to facilitate comprehension.

Notwithstanding the possible implications foreign accent may have for native speakers (NS), in a global world where English acts as *lingua franca*, it is unreasonable to expect foreign language learners to attain native-like proficiency – especially in pronunciation, given the challenging nature of this competence (see section 1.1.2 below). Furthermore, given that a great deal of interactions in which non-native speakers (NNS) are involved may be with other international speakers, native standards may not be the most adequate or perhaps not even necessary (Jenkins, 2000, 2002; Setter & Jenkins, 2005). Additionally, as noted above, it is perfectly possible to have a foreign accent and still be impeccably intelligible.

---

[3] For a review of the different types of discrimination that may by triggered because of accent see Munro (2003) or Derwing and Munro (2009).

Research has shown that intelligibility and comprehensibility are somehow related, but the same does not hold for accentedness, which has been found to be a rather bad predictor of intelligibility (Munro & Derwing, 1995; Derwing & Munro, 1997).[4] The findings by Munro & Derwing show that stimuli that are perceived as highly accented can be nonetheless perfectly transcribed by judges (i.e. they are perfectly intelligible). They report that samples of speech considered to be highly intelligible are not necessarily judged as little accented (Munro & Derwing, 1995). Thus, even though being little intelligible may correlate with a strong foreign accent, the opposite is not always true. In fact, some of the studies that report negative effects of foreign accented speech suggest that these negative associations ultimately stem from processing difficulty (i.e. comprehensibility), rather than accent per se (see e.g. Bresnahan et al., 2002; Lev-Ari & Keysar, 2010).

In light of the above, although native-like proficiency may be a goal for some, such as language teachers who wish to sound like natives in order to offer a native-like model to their students, there is agreement that setting such a difficult goal for the average learner is neither reasonable, nor attainable (Levis, 2005; Morley, 1991; Scovel, 1969, 2000). In fact, this could be one of the reasons leading teachers to the conclusion that it is not worth teaching pronunciation, due to the limited probability of getting close to the native model (Munro & Derwing, 2011). Hence, a reasonable objective for the vast majority is being 'comfortably intelligible' (Abercrombie, 1949). For most FL learners, pronunciation teaching should concentrate on those aspects that interfere with intelligibility and comprehensibility, rather than on the reduction of foreign accent (Munro & Derwing, 1995).

In an attempt to find an appropriate syllabus of English pronunciation for international (FL) learners, some researchers have tried to identify 'core' and 'non-core' features of FL phonology that constitute suitable goals for the needs of international communication (see Jenkins, 2000, 2002). As Gilbert (2008) notes, the goal of pronunciation instruction should not 'amount to mastery of a list of sounds or isolated words, but to learning the […] way of making a speaker's thoughts easy to follow' (p. 1). Gilbert (2010) claims that learners should be taught the core elements of spoken English and puts forward a shift in mentality, by focusing on 'accent addition' rather than 'accent reduction'. This, she claims, is 'rather different from the 'stain-removal' approach' (n.p.).

[4] Intelligibility is defined as the extent to which a listener understands an utterance and comprehensibility is understood as the easiness with which the listener is able to understand the message (i.e. processing difficulty) – see Munro and Derwing (1995) and Derwing and Munro (1997).

Because a great deal of research has focused exclusively on what is considered intelligible for native listeners, Jenkins' proposal contemplates the fact that features that may hamper intelligibility for native speakers (NSs) may not be the same that pose problems for NNSs. One of Jenkins' observations is that in the case of English, native speakers represent a small minority of speakers throughout the world, questioning whether having native-like proficiency as the goal is appropriate. As Jenkins (2002) puts it, 'regional variation is the (acceptable) rule rather than the (unacceptable) exception' (p. 85).[5]

Research has shown that students tend to have less favourable attitudes towards certain accents when they perceive them as difficult (Cenoz & García Lecumberri, 1999a). In this regard, following Cook's (1999) idea that using intelligible and comprehensible models of non-native pronunciation as 'legitimate aspirational models' may be more productive, Murphy (2014) advocates the use of intelligible non-native models of pronunciation. He claims that since few FL learners will ever become native-like accented English speakers, using non-native models may produce results that are satisfactory enough in terms of effective communication. This should allow learners to understand that being intelligible is possible and that sounding non-native is not bad. Murphy (2014, p. 261) notes that '[al]though the inclusion of non-native speech samples may seem controversial at first, it is a characteristic rich in potential for illustrating to learners that the pronunciation models and goals being presented are feasibly within their reach'.[6] He recommends encouraging learners to identify what intelligible non-native speakers are doing well, rather than perpetuating a deficit model that focuses on non-native difficulties. In fact, even students receiving pronunciation training have requested non-native models at beginner levels in order to be able to imitate them, given the encountered difficulty in imitating native speakers (Pi-Hua, 2006). Moreover, this could also help increase non-native language teachers' self esteem, as the vast majority of foreign language teachers around the world are competent non-native speakers.

1.1.1 Features that need to be addressed when teaching pronunciation

Related to pronunciation goals are the aspects of pronunciation that need to be addressed, for prioritising certain features over others will affect the final outcome of pronunciation instruction. Researchers have long advocated a more 'communicative pronunciation'

---

[5] She exemplifies her claim with Britain's Received Pronunciation (RP), which is used by less than 3% of British speakers (Crystal, 1995, p. 365 – as cited in Jenkins, 2002, p. 84).

[6] The example offered by Murphy (2014) in the study is Javier Bardem's speech, someone who is considered to have a marked foreign accent and yet is perfectly comprehensible.

(Morley, 1994), with a shift away from the traditional drilling of isolated sounds and stressing the importance of suprasegmentals and communicative activities that enable learners to become confident and intelligible speakers (e.g. Celce-Murcia, 1987; Celce-Murcia, et al., 2010; Derwing, Munro, & Wiebe, 1998; Gilbert, 1994, 2008; Levis, 2005; Munro & Derwing, 1995; Morley, 1991, 1994; Pennington, 1989, 1999; Pennington & Richards, 1986). In this respect, research has offered evidence in favour of teaching both segmental and suprasegmental features.

As far as segmentals are concerned, several studies indicate that errors at the segmental level are correlated with ratings of accentedness (Derwing et al., 1998) and measures of intelligibility and comprehensibility (Derwing et al., 2000). Some studies emphasise the importance of individual sounds in stressed positions (Zielinski, 2008), while others suggest that it is vowels that most affect intelligibility (Bent, Bradlow, & Smith, 2007; Schairer, 1992). Segmental mispronunciations have been judged more harshly than intonation in terms of irritability of the listener (Fayer & Krasinski, 1987), but more importantly, segmental mispronunciations may also pose problems at the prosodic level. For example, inadequate tempo durations have been ascribed to lack of fluency when articulating segmental sound sequences (Gutiérrez-Díez, 2001). When learners cannot control segmental articulation adequately, it takes longer for them to articulate words, consequently affecting issues such as tempo, rhythm, and probably intonation. Hardison (2004) also reports segmental problems affecting suprasegmental features in her pre-test, although these were palliated after administering training of suprasegmentals. Finally, Albrechtsen, Henriksen, and Færch (1980) found that learners with few segmental errors usually showed good intonation and little hesitation.

Similarly, it could be hypothesised that an inadequate pronunciation of schwa as a strong, full vowel, could directly affect suprasegmental features given its fundamental role in the stress-timed nature of English rhythm. Misprounouncing this vowel as a full stressed vowel can alter the pronunciation of the whole word, affecting not only the word itself, but also the prominence received by the word in the wider context of the sentence, and consequently, also suprasegmental phenomena such as rhythm or intonation. The pronunciation of schwa as a full vowel is one of the most common problems for Spanish learners of English (see Monroy, 2001); by pronouncing common English weak forms (e.g., *the, a, of, that*) as strong, every word would be perceived as accented, therefore seriously affecting the English rhythmic structure.

As regards suprasegmentals, they have been claimed to be 'the backbone of speech, providing the structure that links the individual sounds to one another and to the linguistic substrate' (Eskenazi, 2009, p. 837). Prosody is fundamental for listening comprehension, as it helps organise information and 'guide' listeners (Gilbert, 2008). Stress at the sentence level guides listeners by marking the most important information; it works as some kind of 'telegram', where the most important parts of the message are made prominent and words that are not important (e.g. function words) are weakened (Avery & Ehrlich, 1992). Gilbert (2010) goes as far as to claim that no matter how much learners practise individual sounds, they will not improve their intelligibility without the appropriate mastery of prosody. Others sustain that an adequate prosody can 'offset mediocre articulation' (Eskenazi, 2009, p. 837). Gilbert (2008) reports that learners often complain about native speakers talking too fast, something she ascribes to the learners' inability to process speech adequately. Something as common as weak forms, contractions, sound assimilation or linking processes of connected speech will dramatically alter words as students know them (i.e. their citation form).

If we look at the research in favour of prioritising suprasegmentals, Anderson-Hsieh, Johnson, and Koehler (1992) compared the effects of deviance in segmentals, prosody and syllable structure on intelligibility with samples taken from speakers of 11 language groups and found that prosody was the aspect that exerted the strongest impact on intelligibility. Research has shown that lexical stress plays a vital role in intelligibility both for native and non-native speakers (Field, 2005). Misplacement of nuclear stress has been found to hamper comprehension and affect the listeners' impression of the speaker's communicative abilities (Hahn, 2004). Adequate use of pauses has also been shown to affect judgements of intelligibility (Fayer & Krasinski, 1995) and comprehension (Blau, 1990). In a similar line, although the group receiving training in segmental accuracy in the study by Derwing et al. (1998) obtained the best results in terms of accentedness, it was the group receiving suprasegmental training that could transfer improvements to more extemporaneous tasks and the one that improved most in terms of comprehensibility and intelligibility.

Because research findings offer support in favour of both approaches, researchers have come to the conclusion that teachers should seek a balance between segmental and suprasegmental features (Derwing et al., 1998). Efforts should be directed at improving those aspects which directly affect intelligibility and comprehensibility – arguably the most adequate goal given the very limited time pronunciation receives in general English as a foreign language (EFL) classes (Derwing & Munro, 2009).

There is ample research investigating how particular aspects of pronunciation affect intelligibility; however, research is needed to shed light on the aspects that are most important for different L1s. Research should not only address the way native speakers perceive the pronunciation of foreigners, but also how their pronunciation is perceived by speakers of different L1s. As Jenkins (2002) notes, we cannot assume that native-speaker measures of intelligibility will be the same as for other L1 backgrounds. The linguistic distance between phonological systems of different languages may yield some aspects of FL speech unintelligible for one particular language group, but not for another.

Authors like Field (2005) have claimed that in the 'intelligibility debate, […] the occasional insertion of a nonstandard phoneme should not grossly disrupt communication' (p. 402), with data offering support for his claim. However, counterevidence can be found in a study he cites by Cutler and Clifton (1984), where stressed syllables were switched in a number of words in order to test their impact on word identification. The authors found that intelligibility was most compromised when there was not only a shift in stress, but also a change of vowel quality, as in *lagoon* [ləˈguːn] being mispronounced as [ˈlægun], or wallet [ˈwɒlɪt] mispronounced as [wɒˈlet] (cited from Field, 2005, p. 404). Something Field does not seem to take into account is the fact that when a non-native speaker substitutes one vowel for another, it will not always confirm native-speaker predictions. The first instance of a Spanish mispronunciation that comes to mind when reading these examples is the wrong pronunciation of *comfortable* as *[komforˈteibol],[7] where there is not one vowel change, but *four* changes that completely distort its lexical stress patterns due to the pronunciation of schwa as a full Spanish vowel (see Monroy, 2001 for examples).

Some EFL learners are strongly influenced by orthography, as exemplified in common mispronunciations where learners pronounce silent letters (as the <l> in *half* or the <s> in *aisle*), substitute phonemes due to an overgeneralisation of common sound-spelling associations in the FL (as the <ch> in *archives* pronounced as /tʃ/, or the <th> in *worthy* pronounced as /θ/), or due to unusual grapheme-phoneme correspondences (as the <ea> in *steak*, commonly mispronounced with /iː/).[8] When considering examples like *comfortable* *[komforˈteibol], it is difficult not to assume that recurrent pronunciations of this kind will pose high comprehensibility demands on listeners. This should have a strong

[7] An asterisk (*) is used before phonemic or phonetic trasncriptions in this dissertation in order to represent interlanguage forms or mispronunciations.
[8] For more examples, see Mompean and Fouz-González, in press; Monroy, 2001; or chapter VI in this dissertation.

impact on both intelligibility and comprehensibility. Something similar will happen with the Spanish pronunciation of the word *Georgia* with a complete transfer from the Spanish pronunciation /xeˈorxja/; the Spanish version is nothing like the English one /ˈdʒɔːdʒə/ (not only in terms of vowels, but also in terms of consonants), and correct identification seems extremely unlikely for listeners with different phonological systems, or with systems with different sound-spelling correspondences. In fact, Field (2005, p. 412) reports that 'a curious finding' for the group of Spaniards was that shifting stress and changing vowel quality actually improved intelligibility. Given the examples of vowel substitutions by Spanish speakers offered above, this is not surprising. Walker (2005) has exemplified that when two interlocutors share a common L1, they may resort to L1 sounds to resolve possible ambiguities. The example he gives is *wrote* versus *road*, where students tried to clarify the latter by pronouncing it as /ro-at/ (p. 551). In these cases, learners will try replacing FL sounds with L1 sounds or, in the case of Spanish, also doing a 'phonetic reading' of the orthographic representation. This type of segmental substitutions /rəʊd/ (*road*) for /roat/ will undoubtedly make identification extremely difficult (if not impossible) for FL learners not sharing a common L1.

The above discussion suggests that we cannot deny the importance of segmentals or suprasegmentals, but try to find a balance between both, always with intelligibility and comprehensibility (in that order) in mind if there are time constraints that force us to prioritise one or the other.[9] Derwing et al. (1998) found that students who received training on segmental aspects improved considerably in terms of accentedness in those cases where they were able to monitor their production, whereas this improvement did not transfer to more spontaneous tasks. Nonetheless, learners who received training in rhythm, intonation and stress were able to transfer improvements to extemporaneous production. They argue that a balanced approach is in order, given that while global instruction may help learners in terms of spontaneous tasks, segmental training is useful to self-correct mispronunciations that may cause breakdowns in communication. This is in line with Morley's (1991) proposal of a dual-focus approach combining a microlevel (bottom-up) and a macrolevel (top-down) focus, where either of the two can be prioritised depending on the learners' needs at a particular time.

Authors like Brown (1991) or Catford (1987) have proposed ranking the saliency of errors on the grounds of their functional load, that is, 'a measure of the work which two

[9] See MacCarthy (1978) for a practical classification of pronunciation error types.

phonemes (or a distinctive feature) do in keeping utterances apart' (King, 1967, p. 831).[10] However, only the study by Munro and Derwing (2006) has tested the hypothesis empirically. The hypothesis was tentatively confirmed in that high functional load errors showed stronger correlations with accentedness and comprehensibility than low functional load errors. Moreover, in line with their previous findings, low functional errors were also perceived by listeners, as reflected in their ratings of accentedness, but these errors did not exert such a strong influence on comprehensibility (therefore suggesting differences among the three constructs – accentedness, comprehensibility and intelligibility). Nevertheless, more research is needed in order to support the functional load principle empirically. As Munro and Derwing (2006) note, while they focused on the role of consonants, other aspects of speech should be addressed too, including vowels and suprasegmental phenomena. In addition, they also recommend using speech in controlled and natural interactions, as well as listeners from native and non-native backgrounds.

1.1.2 The challenge of pronunciation

Pronunciation is one of the most challenging competences to acquire in a foreign language. It is the only competence in which learners very rarely (if ever) acquire native-like proficiency (see e.g. Bongaerts, van Summeren, Planken, & Schils, 1997; Flege, Munro, & MacKay, 1995; Moyer, 1999). Even though native-like performance is seldom the goal for FL learners (see below), the fact that pronunciation is the only competence that cannot be (or is not often) mastered entirely in a foreign language is a clear indicator of its complexity. In other aspects of language, like grammar or lexis, learners can attain native-like mastery with certain guarantee of success. However, the acquisition of pronunciation requires psycho-motoric and perceptual skills, which are besides further constrained by affective factors.

On the one hand, learners need to be able to *perceive* the FL phonology properly, which does not depend on their effort or declarative knowledge, but on their perceptual ability (i.e. their auditory capacity to perceive fine distinctions between similar sounds and other suprasegmental phenomena, as well as their ability to categorise them accurately). And on the other hand, they need to be able to *produce* the FL sounds and patters adequately, which requires the adequate performance of a series of articulatory movements they may not be accustomed to (i.e. modification of the already-established patterns of

[10] Cited in Munro and Derwing (2006, p. 522).

articulation that they use in their L1). There are many learners who, after years of study and practice, and even after spending several years of their lives abroad, continue to mispronounce the foreign language. In some cases, due to an inability to perceive or conceptualise the FL sounds and patterns correctly; in others, because they are unable to perform the necessary movements to articulate them.

Furthermore, apart from the cognitive, psychomotoric and perceptual, there are also a number of psychological factors that interfere with the acquisition of pronunciation (e.g., shyness, ego permeability). Some of these factors have an influence on the rest of competences too, but they affect pronunciation to a much greater extent. In addition, mistakes in pronunciation are much more noticeable than in other aspects of language, as they become apparent the moment speakers open their mouths. When evaluating someone's mastery of grammar or lexis orally, a listener may condone certain mistakes if the speaker corrects them immediately. These can be considered as slips of the tongue, performance mistakes that even native speakers frequently make (Corder, 1967). However, as far as pronunciation is concerned, the very moment something is mispronounced, the listener immediately realises that (s)he is not talking to a native speaker.

One of the strongest determinants in the acquisition of FL phonology is the learners' first language. Researchers have long identified the learners' L1 as the main predictor of their behaviour in the L2 (Lado, 1957; Wardhaugh, 1970; Weinreich, 1953). A clear indicator of the influence exerted by the mother tongue is the fact that it is usually very easy to spot people's nationality because of traits of their L1 that are present in their accent. As Collins and Mees (2009, p. 210) put it, 'although not every pronunciation problem can be traced back to the sound system of the learners' L1, it is no exaggeration to say that most of them can indeed be predicted from contrastive analysis'. The mother tongue may influence acquisition from the point of view of perception, by acting as some sort of 'phonological sieve' (Trubetzkoy, 1958/1939)[11] that filters out every aspect of L2 phonology that is not relevant in the L1, but also from the point of view of production, as learners need to establish new articulatory habits that may be different to the ones they use in their L1 (Flege, 1987).

Numerous researchers assume that a good part of learners' errors stem from an inadequate perception of the FL phonology, which is strongly conditioned by the phonological system of the learner's first language (see e.g. Best, 1995; Best & Tyler, 2007;

[11] Cited in Monroy (2001).

Flege, 1991, 1995; Kuhl & Iverson, 1995). These theories postulate that if learners hear a sound that is similar to a sound they already have in their L1, they will assimilate that sound to the same phonetic category they have for their native sound. It is assumed that if learners do not have 'accurate perceptual 'targets' to guide the sensorimotor learning of […] sounds, productions of the L2 sounds will be inaccurate' (Flege, 1995, p. 238), as they will resort to the same articulatory movements that they use for the articulation of L1 sounds (Flege, 1987). However, research has shown that not every deviation from L2 phonology can be traced back to the learners' L1, as learners do not always rely on the same perceptual cues when identifying sounds in the L2 and the L1 (Bohn, 1995), and also because of the existence of errors that are developmental in nature (see Major, 1987).

Another challenge FL learners face is age. The age at which learners start to learn the foreign language is considered to determine their success at mastering the FL phonology. This hotly debated issue has often revolved around the Critical Period Hypothesis (Lenneberg, 1967), a theory stating that there is a period between 5 and 11 years of age that presents the most favourable conditions for language acquisition. The hypothesis has found support in cases like Genie, the American feral child deprived of speech until the age of thirteen; although in the case of second and foreign language learning, opinions differ. Some have gone as far as to suggest that it will be almost impossible to acquire the phonology of the second language after such critical period (Scovel, 1969), while others adopt a milder position and argue that the younger a person starts learning the pronunciation of the foreign language, the easier will be for this person to acquire it (Singleton, 1995, 2003).[12]

Several studies offer data that support the view that late learners receiving instruction may pass for native speakers phonologically (Bongaerts et al., 1995). These studies report on students who manage to *perform like* native speakers in certain tasks (e.g. Bongaerts, 1999; Bongaerts et al. 1995; Ioup, Boutsagi, El Tigi, & Moselle, 1994; Moyer, 1999). Hence, Bongaerts *et al.* (1995) advocate replacing the term *critical period* for *sensitive period*, given that the former completely leaves out the possibility to attain native-like pronunciation, while the latter neither rules out this possibility, nor does it deny possible advantages of an early start.

---

[12] See Piske, MacKay and Flege (2001) for a comprehensive review of factors affecting degree of foreign accent and native-like attainment.

Despite the fact that some learners have been able to perform at a native-like level, native-like attainment of pronunciation can be considered to be quite rare (Bongaerts et al. 1997; Moyer, 1999). Moreover, as Major and Kim (1999) point out, research has yet to prove that these learners who perform 'like natives' can do so 100% of the time. They maintain that even native speakers are sometimes not rated as 'native' in certain tasks. When investigating the validity of the Critical Period Hypothesis, it is important to clarify what is meant by 'native-like pronunciation'. In its most literal sense, it should mean the ability to speak like a native speaker in any context under any circumstances. If you acquire 'native-like' pronunciation, you speak 'like a native speaker', and that should apply to every context. However, when dealing with foreign language learners, being able to imitate the foreign language sounds and patterns like a native speaker would seem quite a satisfactory achievement to many – even if they are not able to sound like a native speaker *all the time* (e.g. after a long day at work, or in situations of stress). Being able to imitate sounds and patterns *like* a native speaker implies that the learner has overcome the perceptual and productive obstacles mentioned above. To a certain extent, that could be interpreted as an indicator that person *is* able to acquire foreign language phonology, even after puberty. It is true that native-like performance would require even more, but at least we could maintain that the learners' abilities to perceive and produce the language remain intact.

The context in which the language is learnt will determine the amount and type of exposure to the target language, but also the learners' opportunities for practice, which is pivotal in the acquisition of foreign language phonology. The amount of comprehensible input learners receive has long been identified as one of the essential ingredients for language acquisition to take place (Krashen, 1982, 1985). This also applies to pronunciation acquisition, as the production of speech sounds is thought to be guided by perceptual targets that develop from the many different instantiations an individual hears of a certain sound (Flege, 1987, 1991). In this sense, children spend their first years of life merely listening to their parents and people around them speaking in their language. They are constantly 'absorbing' and assimilating examples of language, hearing (and unconsciously acquiring) grammatical structures while watching their parents use words to refer to certain objects. Similarly, they are constantly listening to the way people pronounce the words of that language, how they stress them, or the rhythmic and intonational patterns that they use. However, in the same way Krashen claims that input has to be *comprehensible* so that intake takes place, learners in a FL environment have to be able to hear the sounds and patterns of the foreign language *properly* before they are able to pronounce them.

Naturalistic approaches consider language exposure enough for the acquisition of foreign language phonology. They postulate that students will implicitly learn how to pronounce words by mere contact with the language, by imitation. As stated in one of the principles of the Direct Method: 'The child will acquire an excellent pronunciation, getting to perfectly imitate the speech spoken in his environment' (Sánchez, 2009, p. 51). Nevertheless, while this may be true for children learning their mother tongue, surrounded by samples of language twenty-four hours a day, seven days a week, and with no other language interfering, adults learning a foreign language will see their acquisition impeded by the above-mentioned perceptual and productive constraints. Given the role of perception in imitation, as noted above, everything learners hear is influenced by their mother tongue, the filter through which they perceive foreign language sounds and patterns. Consequently, if they are not able to hear the FL sounds and patterns properly, they will not be able to produce them correctly either. By just listening to the foreign language, most adult learners will not be able to 'pick up' a good pronunciation, especially if they have already fossilised certain mistakes. Indeed, in most cases they will not be even able to perceive the difference between what they pronounce and the target they are supposed to attain.

Whether the language is learnt as a second or foreign language will influence acquisition significantly. Children learning their native language have infinite samples of language at hand, whereas adults learning a foreign language will need to make an effort in order to find this input somewhere. If learnt as a foreign language, exposure will commonly be limited to a few hours a week and only in a school context, 'with non-native teachers and no communicative need to use the foreign language outside the classroom' (Cenoz, 2003, p. 78). In foreign language contexts, instead of presupposing an inherent (in)ability on the part of young learners to attain native-like pronunciation, the role age will play will be a consequence of other variables, like the already-established patterns of the mother tongue, the amount of exposure to the foreign language and opportunities to practice, or the amount of prior instruction. This claim finds support in studies like the one by García Lecumberri and Gallardo (2003). They compared the perception and production of three groups of children of different ages (4, 8 and 11 years old) and found that older students were judged as having a more target-like accent and as more intelligible.[13] Thus, their data

[13] Similarly, acquisition of a new phonological system in L2 contexts may be more difficult for some than for learners in FL contexts, since adults in L2 contexts will not normally receive explicit instruction (Leather, 2003). Hence, they will be subject to the same perceptual constraints imposed by their L1 without the necessary training to overcome them. Furthermore, they input learners in L2 contexts are exposed to is not as

suggest that an early starting age is not a facilitating factor in non-natural exposure, at least in the medium term. Hence, despite the fact that age-related pronunciation problems were originally ascribed to biological age, they seem to be more related to FL experience and context where the FL is learnt, length of residence in or age of arrival to the foreign country, or quality and quantity of input (see Flege, 2009).

Notwithstanding the above, even if learners in FL contexts manage to obtain the necessary amount of comprehensible input (and, in the case of pronunciation, perceive the FL phonology properly), input alone will still not be enough, as they will also need to be able to adequately perform the articulatory movements required by the FL phonology. Inspired by the failure of French immersion programmes in Canada in terms of learners' mastery of productive skills (speaking and writing), which were much more limited than their receptive skills (i.e. listening and reading comprehension), Swain (1985) identified the missing ingredient in Krashen's recipe for language learning success: output. Learners in those programmes were exposed to massive amounts of (comprehensible) input on a daily basis, and yet, they did not show the same levels of mastery as native speakers.

By producing language (i.e. output), learners are able to notice gaps in their knowledge as they try to express something and they realise they do not know how, they can test their hypothesis about the way language works thanks to their listeners' reactions (Long 1996), and they can also use this language to reflect on the language itself (Swain, 1985, 2005). In the case of pronunciation, producing language will also be a way of noticing gaps in the sudents' productive abilities, testing their hypotheses and reflecting on their pronunciation (see section 1.1.5). Even though research has shown that improvements in perception can lead to improvements in production (e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Ramírez-Verdugo, 2006; Thomson, 2011; Wong, 2015), output also plays a vital role in the acquisition of pronunciation, as learners need to automatise new articulatory habits so that they can produce FL sounds and patterns naturally and effortlessly.

Affective factors, such as shyness, anxiety, motivation or attitude play also a role in matters of pronunciation (Skehan, 1989). Pronunciation has been considered to be 'the most intimate component of linguistic identity' (Cortés Moreno, 2000, p. 97). The way we speak conveys a great deal of information about ourselves (Setter & Jenkins, 2005), and

---

'native-like' as that of learners acquiring their first language, as they will often be exposed to the foreign accented speech of other immigrants, which may even reinforce their foreign accent (Flege, 2009).

changing our accent or our way of speaking when using a foreign language is something that inhibits learners from practising freely. When two languages are phonetically distant, learners may feel weird or shy when they need to imitate sounds they may consider ‘funny’. For example, Spanish learners of English usually feel funny when imitating the VOT values of English alveolar plosives; they perceive this pronunciation as ‘affected’. Thus, it could be hypothesised that the more different the FL and the L1 sounds are, the shyer students might feel. As a case in point, the researcher’s students very often mispronounce words on purpose when using English words with their Spanish friends, even if they are all studying English at University level. This conscious mispronunciation suggest that learners feel unnatural if they have to carefully pronounce a foreign word in the middle of a conversation in their L1.

The above-mentioned examples are related to what Guiora, Beit-Hallahmi, Brannon, Dull, and Scovel (1972) describe as *language ego*, ‘a maturational concept [that] refers to the self-representation with physical outlines and firm boundaries’ (p. 421). They claim that pronunciation is the most salient aspect of language ego and the hardest to penetrate, which could be compared to taking on a new identity. Moreover, they state that in the same way that it is the hardest aspect to penetrate when acquiring a new language, it is also the most difficult to lose in the first language. They claim that language ego is rather malleable in the early stages of development. In their view, this could explain why it is easier for children to imitate FL sounds with precision, whereas it is more difficult (or perhaps somewhat ‘embarrassing’) for adults. The hypothesis has found support in studies showing that learners obtained better results in a number of pronunciation tasks when being less self-aware, be it under the effects of hypnosis (Schumann, Holroyd, Campbell, & Ward, 1978), alcohol (Guiora et al., 1972), or alcohol and tranquilisers (Guiora, Acton, Erard, & Strickland, 1980). Needless to say is the fact that shyness is extremely counterproductive, as learners will not take advantage of their opportunities to practise with other speakers (Morley, 1994) and automatise correct articulatory habits. When learners gain confidence, they feel in a better position to practise and monitor their pronunciation (Hardison, 2004; Ramírez-Verdugo, 2006; Smith & Becknamm, 2010).

Another factor that may seriously affect pronunciation is anxiety. Even though experiencing anxiety has long been considered to affect language acquisition (e.g. Horwitz, Horwitz, & Cope, 1986; Krashen, 1985), pronunciation anxiety has been described as the aspect that causes learners the greatest fear when tackling the FL (Baran-Łucarz, 2014). Baran-Łucarz (2014, p. 38) defines pronunciation anxiety as ‘a feeling of apprehension experienced by FL learners either in the FL classroom or natural setting, deriving from negative FL pronunciation self-perceptions, fear of negative evaluations, and beliefs about the importance of pronunciation, difficulty of learning and the sound of the FL pronunciation, evidenced by typical cognitive, physiological/somatic and behavioral symptoms of being anxious'.

Pronunciation anxiety can be manifested in either a poorer performance than usual, or outright reluctance to speak in the FL. In the researcher's informal observations, students' pronunciation may range from outstanding to average depending on the stress of the situation. Learners may have been pronouncing perfectly well for a couple of minutes and then their performance decreases considerably once they have made a mistake and have become self-conscious about their pronunciation. As a case in point, participans in Derwing and Rossiter's (2002) study reported awareness of changing their accent when they were excited or nervous. As regards the learners' willingness to communicate, Baran-Łucarz (2014) found that it can be affected considerably by pronunciation anxiety, especially when talking to people who are familiar to the learners (e.g. friends, classmates), as learners are afraid of 'losing face'. Furthermore, students' self-perceptions of their pronunciation have been shown to be worse under high levels of anxiety than under low levels of anxiety, which may limit their willingness for subsequent interactions in the FL (Szyszka, 2011).

An additional factor that will contribute to the learners' success in acquiring the FL pronunciation is the attitude they have towards learning the FL pronunciation, which is closely linked to their motivation. Pennington (1994) has stressed the learners' willingness to change as well as the perceived usefulness to change their pronunciation as essential factors that will determine ultimate success. As regards the former, given the extreme difficulty of acquiring native-like pronunciation, it seems safe to assume that only highly motivated learners will be the ones who make the necessary effort to attain it. As for the latter, learners will not try to modify their pronunciation if they believe they will be understood anyway (e.g. a sloppy pronunciation of *hello* as *[ˈxelou]). As a case in point, several studies have found significant correlations between the degree of concern about having a good pronunciation and performance (e.g. Elliott, 1995; Suter, 1976; Purcell & Suter, 1980). Similarly, the participants who were highly proficient in the studies by Bongaerts et al. (1997) and Moyer (1999) were language teachers and university level teachers, therefore a clear example of learners who have a very strong motivation for (and consequently a very positive attitude towards) learning the language. It is worth noting, though, that in Elliott's (1995) study, despite finding correlations between learners' positive attitudes and pronunciation performance, those learners with more positive attitudes were not the ones that most benefited from instruction.

One final aspect that will condition pronunciation acquisition is aptitude. It has to do with the individual characteristics and abilities of the learner and it will affect pronunciation in the same way as any other skill in life. In the same way that some people are better than others at maths, or faster runners than others, it seems logical to hypothesise that some people may have a special ability for learning languages, and in particular, to acquire the pronunciation of a foreign language. As defined by Carroll (1981):

> Aptitude as a concept corresponds to the notion that in approaching a particular learning task or program, the individual may be thought of as possessing some current state o capability of learning that task – if the individual is motivated, and has the opportunity to do so. That capability is presumed to depend on some combination of more or less enduring characteristics of the individual […] Even if no measures of foreign language aptitude were available, there would be general agreement that individuals appear to differ in their capability of learning foreign languages. On the one hand, some individuals seem to be able to acquire foreign languages easily and quickly, even when not particularly well motivated to do so. On the other hand, some individuals have marked difficulty in acquiring a foreign language, even when highly motivated and interested in doing so (pp. 84-85).

Carroll (1962) proposes four major components language aptitude is composed of: phonetic coding ability, grammatical sensitivity, rote learning ability for foreign language materials, and inductive language learning ability.[14] As regards phonetic coding ability, our main concern here, he claims that it is 'an ability to identify distinct sounds, to form associations between those sounds and symbols representing them, and to retain these associations' (Carroll, 1981, p. 105). Even though some have claimed that 'much of what is considered second or foreign language aptitude [is] directly related to conscious learning' (Krashen, 1981, p. 158), one could hypothesise that in terms of pronunciation acquisition, aptitude goes beyond cognitive capacities, as some people may have a 'better ear' than others (e.g. the ability to hear and imitate voices or accents). In any case, as Piske et al. (2001) note, more research is needed before solid statements can be made, as there is a lack of empirical research investigating the different factors that contribute to language aptitude in a controlled manner.[15]

---

[14] Cited in Carroll (1981, p. 105).

[15] The reader is referred to Piske et al. (2001, p. 202) for a brief review of studies that have investigated aspects such as musical ability or mimicry ability in relation to language aptitude.

1.1.3 Pronunciation: the neglected competence

Despite the vital importance of pronunciation for communication and the problems that may arise from inadequate pronunciation, the teaching of this competence is very often neglected in FL teaching. As Setter (2008) puts it:

> Like listening, pronunciation is sometimes neglected in the process of language teaching in favour of reading and writing, which are rather more likely to lead to success in examination in cultures which are so much more focused on those media. However, it seems rather pointless to study a (living) foreign language at all if one does not intend to communicate in it with other speakers of that language and, to this end, one must learn how to pronounce it in a way which can be understood by a variety of listeners. English […] has arisen as a world *lingua franca*, and being intelligible when speaking it has therefore received a great deal of attention (p. 447).

One of the reasons why pronunciation has not received the attention it deserves is because the interaction of the sound system with meaning has often been ignored (Celce-Murcia et al., 2010). Since the arrival of CLT, the primary goal in FL teaching/learning has been to develop the learners' communicative competence, placing a higher value on fluency and appropriate communication, and therefore drawing attention away from pronunciation – commonly perceived as more related to accuracy and linguistic competence than to communicative competence (Pennington & Richards, 1986) and as having limited value for communication (Cenoz & García Lecumberri, 1999a).

As Morley (1994, p. 66) notes, 'conventional wisdom has maintained that pronunciation is not important', that 'students will pick it up on their own', or that it cannot be taught, even if teachers wanted to. Indeed, the fact that teachers feel inadequately prepared to teach pronunciation is one of the most common claims in the literature (Burgess & Spencer, 2000; Breitkreutz, Derwing, & Rossiter, 2001; Henderson et al., 2015; MacDonald, 2002). This leads to unconfident teachers who teach it infrequently and unsystematically (Levis, 2007), often addressing it as a reactive response to problems that arise, rather than being strategically planned (Kelly, 2000).

Emphasis is often given to other skills as a result of limitations of class time (Levis, 2007; Setter, 2008) or due to a shortage of adequate materials (Busà, 2008; Calvo-Benzies, 2013) – something that is further aggravated by the fact that teachers' lack of training in pronunciation renders them unable to select those that might be suitable (Derwing & Rossiter, 2002). Additionally, some have pointed out the fact that it may not be appropriately emphasised in curricula (Busà, 2008), although, even in those cases where the law does reflect pronunciation goals, there is often a considerable disparity between curricula and legislation and what actually happens in class (Calvo-Benzies, 2013; de Jódar, 2008). The fact that pronunciation is usually not assessed is also important (Henderson et al., 2015), given that when something is not assessed, it is bound to be neglected. In any case, the idea that pronunciation is poorly handled in many language programs is fairly widespread.

Pronunciation has also been marginalised in terms of research. Derwing and Munro (2005) cite a number of works in applied linguistics as well as teacher preparation books where pronunciation is either not at all mentioned, or treated superficially with no reference to empirical research. Furthermore, this neglect is also perceived by students, who often express their desire to devote more time to learning pronunciation, a competence they perceive as difficult but important for communication (Calvo-Benzies, 2013a). Learners mention having frequent requests for repetition and consider pronunciation as one of the primary causes for their language/communication problems (Derwing & Rossiter, 2002). In this regard, learners' views are relevant not just because they offer information about them, but because 'their beliefs and attitudes are very likely to influence the second language learning process' (Cenoz & García-Lecumberri, 1999b, p. 5).

If we focus on the particular context of Spain, Calvo-Benzies (2013b) conducted a study with 138 participants enrolled in secondary education in Galicia and found that more than half of the students attended extra-curricular lessons in addition to their regular classes. The reasons for those students to resort to extra tuition were that they wanted to practice spoken English (34%), practice pronunciation (31.1%) and improve grammar (31.9%). This clearly suggests that students do not feel satisfied with what they receive at school, especially in terms of oral skills. The participants in Calvo-Benzies's study were aware of the importance of pronunciation and of the insufficient time devoted to its teaching. They acknowledged that there was a considerable difference in the higher value placed on grammar, reading, writing and vocabulary as compared to speaking and listening.

1.1.4 Can pronunciation be taught?

Some of the above-mentioned reasons why pronunciation has been neglected were the belief that it is acquired naturally and does not need to be taught (e.g. Krashen, 1979) or that instruction will not make a difference (Suter, 1976). Some have gone as far as to suggest that 'pronunciation training is tedious and […] rarely 100 per cent effective', recommending that North American university students be trained to understand foreign accented speech instead of investing money in training teaching assistants (Rubin, 1992, p. 514).

A clear indicator that pronunciation *can* be taught is the fact that, aptitude apart, people who have a degree in Philology (be it English studies, French studies, etc.) usually have a good pronunciation, some near-native or even native-like. Even though some individuals will undoubtedly acquire pronunciation naturally if living in the foreign country for a considerable time, the fact that there are foreign language learners (in FL contexts) who manage to attain a native-like or almost native-like proficiency in terms of pronunciation suggests that training is effective. In fact, after a review of some of the studies claiming that pronunciation instruction does not make a difference, Pennington (1998) concludes that reason may be the nature or amount of training, not an inherent inability on the part of the learners. As a case in point, those studies arguing that native-like pronunciation *can* be acquired after puberty often mention cases in which learners had received extensive training (Bongaerts et al., 1995; Bongaerts et al., 1997; Moyer, 1999).

In any case, the myriad of studies reporting benefits of pronunciation training suggests that 'lack of beneficial effects of training' should never be an excuse to ignore pronunciation teaching (e.g. Cenoz & García Lecumberri, 1999a; Couper, 2003, 2006; Derwing, Munro & Wiebe, 1997; Derwing et al., 1998; Lord, 2005, 2008; Ramírez Verdugo, 2006; Wang & Munro, 2004), with some studies (especially those using technology for pronunciation training) showing considerable benefits in aspects that were challenging for learners after only a couple of weeks, days, or even hours of training (see e.g. Akahane-Yamada, Adachi & Kahawara, 1997; Flege, 1989; Gick, Bernhardt, Bacsfalvi & Wilson, 2008; Molholt, 1988; Saito, 2013; Schwab, Nusbaum & Pisoni, 1985).

Because pronunciation is such a challenging competence, and given the perceptual and productive constraints learners are likely to face due to their first language, tailored instruction aimed at helping them overcome those obstacles will be of great help. As Pennington (1998) recommends, '[i]nstruction can help to provide the learner with the perceptual and the productive experience needed to reconceptualise the targets for phonological performance' (p. 337). By gaining awareness of their learning process and having a clear target model in mind, they could learn to monitor their pronunciation progress (see section 1.1.5). Learners could focus on those problematic aspects that they need to reinforce by trying to imitate them and then move from controlled practice to a more automatised, extemporaneous kind of production (Morley, 1991, 1994).

In light of the above, there is no doubt that something is needed to help teachers handle pronunciation instruction effectively. Despite the fact that those who attain native-like mastery in pronunciation are commonly those people receiving extensive training or doing a degree in English studies, for example, the average FL learner does not have four years to study a degree in modern languages just because (s)he wants to learn the language adequately. We need to find something that allows teachers to implement pronunciation instruction in their FL classes, especially in communicative approaches that sometimes invite neglecting pronunciation in favour of other skills. Given the cognitive, perceptual, psychomotor, and affective factors mentioned above, an avenue for solution is technology, as it offers numerous possibilities to enhance the learners' perception and production of the FL phonology.

1.1.5 On the acquisition of FL pronunciation

As discussed above, numerous researchers identify the learners' L1 as one of the strongest determinants in FL phonological acquisition (Best, 1995; Best & Tyler, 2007; Flege, 1991, 1995; Kuhl, 1991; Kuhl & Iverson, 1995; Lado, 1957). Theoretical models of speech perception postulate that learners' perception of the FL phonology is conditioned by their L1, leading learners to assimilate the sounds and patterns of the FL to the ones they already have for their native language (Kuhl, 1991; Flege, 1987). This, in turn, is manifested in the learners' production of the FL, often showing transfer of phones, allophones and prosodic patterns from their native language (see e.g. Cruz-Ferreira, 1987; Monroy, 2001; Ramírez-Verdugo, 2006). This explains, for example, why Spanish learners of English pronounce English voiceless plosives as unaspirated in initial position, or why they pronounce /t/ and /d/ as dental instead of alveolar. These models do not ascribe *all* pronunciation errors to the learners' L1. However, one of the tenets in this view of speech perception is that if learners do not have 'accurate perceptual "targets" to guide the sensorimotor learning of […] sounds, productions of the L2 sounds will be inaccurate' (Flege, 1995, p. 238), as they will resort to the same articulatory movements that they use for the L1 sounds (Flege, 1987).

Speech perception and production: category and concept formation

The relationship between perception and production is extremely complex and not entirely clear (see Llisterri, 1995 for a review). Numerous studies have offered evidence supporting the view that accurate perception of the FL phonology may precede production (see Flege, 1995). However, research has also shown that accurate production of a problematic sound contrast may precede adequate perception of that contrast (Goto, 1971; Sheldon & Strange, 1982).

The results by Sheldon and Strange (1982) are often cited as evidence that production may precede perception, as Japanese participants in their study were able to mark the distinction between English /r/ and /l/ in their production but were not able to identify these sounds correctly in an auditory task. Nonetheless, the fact that learners' abilities to mark a problematic contrast in production may surpass their capacity to discriminate between two sounds they perceive as similar (i.e. affected by equivalence classification – Flege, 1987) does not necessarily mean that they did not have a prior mental representation of the sounds in question. It is rather bewildering to think that a sound can be attained in production without a mental model whereby the brain activates the necessary articulatory movements in order to produce this sound. For example, if learners of Spanish as an FL have never heard the Spanish trill /r/ because it does not exist in their language, how can they pronounce this sound for the first time? In other words, how can they know what the sound *sounds like* if they have never not heard/perceived it? They might accidentally come up with an /r/-like production, but in order to consciously mark this contrast in a production task, they must have some kind of mental reference or concept (Fraser, 2006) that guides their production. As defined by Fraser (2006, p. 59):

> [C]oncepts group together aspects of reality that are different from some other known aspect of reality. For example, developing the concept BROWN involves understanding what is 'not-brown'.

Humans perceive and group stimuli into categories showing internal structure (Kuhl, 1991), and researchers have argued that the cognitive ability of categorisation could serve to explain some of the issues phonologies have often been concerned with, such as the fact that allophones or different realisations of the same sound can be treated as 'the same' (Mompean, 2002).

In Sheldon and Strange's (1982) study, learners' productions were elicited through reading tasks. If, for example, learners saw the word *right* (as opposed to *light* – /r/ vs. /l/) and were able to articulate /r/ correctly, they must have had a mental (perceptual) model guiding their subsequent articulation of the sound (i.e. the way they *perceive* or *conceptualise* this sound). Learners may have had some kind of concept (or mental representation) of what /r/ and /l/ sound like, including articulatory information as well as common spellings for those sounds. Hence, could articulate those sounds when seeing <r> or <l> in spelling (e.g. *rice* vs. *lice*), but failed to discriminate/identify those sounds in aural perception. Thus, it may be the case that learners in Goto (1971) and Sheldon and Strange's (1982) study had a concept of what /r/ is and what /r/ is not, but nevertheless failed to perceive the distinction auditorily.

The approach adopted in this dissertation subscribes to the view that an adequate perception of the FL is a pre-requisite for productive success. Research has often addressed the acquisition of FL phonology in terms of phonetic categories (e.g. Flege & Eefting, 1988; Flege & Hammond, 1982; Rochdi & Mora, 2012; Thomson, 2011), as phonetic categories are considered to reflect the learners' perceptual and productive abilities in the FL. A great deal of pronunciation acquisition studies have used imitation tasks in order to measure category formation on the grounds that imitation is often regarded as consisting of three distinct processes. These are 'perception of structural properties in the stimuli being imitated, coding and storage in memory, and regeneration in the form of a motoric code suitable for skilled movement' (Flege & Eefting, 1988, p. 730). Nonetheless, learners may be perfectly able to *imitate* a sound or pattern accurately but fail to come up with that sound in spontaneous production. This may be due to lack of familiarity with the word or its phonological composition, or lack of awareness of the relationship between sounds and orthography. For instance, even if learners are able to articulate English /v/ and /f/ correctly, it may be challenging to choose the adequate consonant for words like *Stephen* /v/, *Philip* /f/, *laugh* /f/, or *daughter* (silent).

Improvements in the learners' perception of a problematic sound contrast or pattern does not automatically translate into mastery of this feature in spontaneous production. Being able to adequately articulate features of the FL phonology in controlled environments is paramount, but learners will ultimately need to be able to use these features fluently and effortlessly. In this respect, Morley (1991, 1994) put forward three modes of practice for pronunciation arranged in order of increasing difficulty: *imitative, controlled* and *extemporaneous*. Imitating a sound can be assumed to be easier than producing it in a sentence in which no aural model is provided and where learners have to retrieve the appropriate sound from memory, for example. Similarly, controlled pronunciation of sound in words or sentences that can be rehearsed should be easier than being able to adequately produce that sound in spontaneous production, where the focus may be on the conversation itself and not on form. Even if learners are able to attain a sound in production and know where to pronounce it, it may be the case that they fail to attain an adequate articulation when engaged in extemporaneous exchanges due to various factors, such as the anxiety of the situation, speed or cognitive demands of the conversation, etc.

As a case in point, one of the most challenging aspects to master for EFL learners is the acquisition of English vowels, as evinced in the numerous examples of vowel substitutions in spontaneous samples of the interlanguage of very advanced FL learners (Monroy, 2001). As far as Spanish EFL learners are concerned, English has three vowels that can be assimilated acoustically to the phonetic space of Spanish /a/, namely /æ/, /ʌ/ and /ɑː/ (sometimes including /ə/, depending on the context).[16] However, Spanish learners do not only need to be able to perceive and articulate these vowels correctly, but they also need to know *when* to pronounce them – which is especially challenging given the different spellings for each vowel (e.g. <oo> in *blood*, <u> in *cut*, <ou> in *cousin,* or <o> in *comfortable* for /ʌ/), sometimes even coinciding for different vowels (<a> for *bag* /æ/, *car* /ɑː/, *ago* /ə/).

This problem could be partly explained by making recourse to Cognitive Phonology and the notion of *conceptualisation* suggested above. As Fraser (2001) points out, a great deal of pronunciation problems do not necessarily stem from articulatory ability, but from the learners' incapacity to *conceptualise* sounds adequately. In Fraser's (2001, p. 20) words:

> [T]he problem is not that the person can't physically make the individual sounds, but that they don't ***conceptualise*** the sounds appropriately – discriminate them, organise them in their minds, and manipulate them as required for the sound system of English. For example, nearly all learners who have trouble with the 's/sh' distinction actually use both sounds in their own languages and can produce each of them easily in certain contexts. The problem is that in their languages the sounds are *conceptualised* differently from the way they are in English. Learners need to 'unlearn' the concepts they have held since babyhood for these sounds, and replace them with the similar but different concepts needed to speak English.

This view of FL phonological acquisition implies that learners not only need to be able to perceive and produce the difference between two similar sounds in the FL, but also to create adequate *concepts* or *mental targets* for the FL sounds and patterns. Nonetheless, this idea of *mental categorisation* should not be interpreted as something unrelated to the traditional view of phonetic category formation (reflecting people's ability to perceive and produce sounds – Flege, 1995), but as a facilitating factor in the acquisition of the FL phonology (see e.g. Couper, 2011).

Research has shown that learners acquire new categories for FL sounds after being exposed to numerous instantiations of those sounds (Flege, 1991, 1995). However, these

---

[16] Schwa may be pronounced as four different vowels of the Spanish vowel inventory by Spanish learners of English, depending on the orthography of the word in question. For example, it is often mispronounced as /a/ in words like *servant* or *policeman,* as /e/ in words like *interesting* or *garden*, as /o/ in words like *forget* or *prison,* or as /u/ in words like *awful, or picture* (Monroy, 2001).

phonetic categories must not be something completely abstract and intangible. Using the above example of the pronunciation of /f/, when learners think about the common sound in the words *fly, off, phone,* or *laugh* they must have some kind of mental reference that allows them to conceptualise the sound. It could be argued that acquiring non-native contrasts should be much more difficult without having a mental model. By having a mental category (or concept) of a sound, for example, learners should be better equipped to notice future instantiations of the sound in the input and classify them as belonging to those categories.

Using the same example of /f/, creating an adequate concept for this sound would imply that learners are aware of the existence of the /f/ phoneme. Similarly, if learners have a mental category for /v/, when they hear the pronunciation of *Stephen*, they could associate it to their category for /v/. In the case of /f/, it is easy to think of the letter <f> as some kind of label with which to associate the sounds corresponding to the phoneme. However, this is more challenging for sounds for which learners may not have the appropriate labels, such as vowels. For example, the grapheme <o> can be associated with /ɒ/ in *pot*, with /ɔː/ in *lord,* or with /əʊ/ in *go*. Likewise, the spelling <u> can be associated with /ʊ/ in *put*, with /ʌ/ in *putt*, or with /ɜː/ in *cur*. This becomes even more problematic for different FL sounds that tend to be perceived as belonging to the same category as a result of the L1 perceptual system (Kuhl, 1991; Flege, 1987).

As a case in point, focusing on the problems Spanish EFL learners may encounter when acquiring English vowels, the phonological system of Spanish only has five monophthongs, while the British English system has 12 (Figure 16). For Spaniards with only one vowel sound (/a/) in the same articulatory space where English has three (/æ/, /ʌ/, /ɑː/), it can be extremely difficult to learn to *mentally classify* these sounds when they hear them in running speech. None of these sounds exists in the phonemic inventory of Spanish (unlike /f/ above, for instance), nor is the average learner familiar with these symbols to 'label' (i.e. categorise) them. Hence, even if extensive exposure to these contrasts can help learners eventually discriminate among these sounds, it will be extremely difficult for them to start *conceptualising* the sounds adequately (in terms of common spellings or contexts of occurrence) without explicit instruction. If learners do not have any referent for these vowel sounds, they will classify instantiations of these three sounds under the most similar category they have in their native language, namely /a/. One way of facilitating this conceptualisation is by familiarising learners with phonemic symbols.

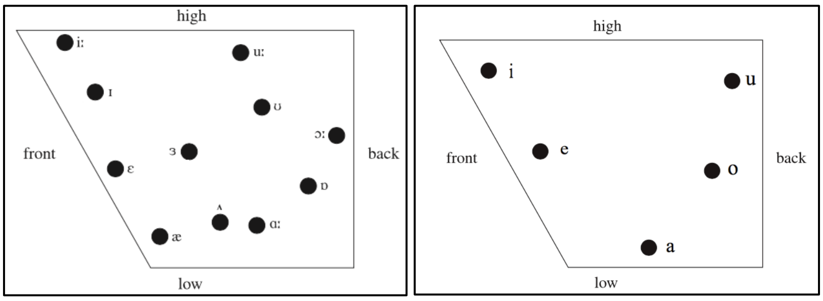


Figure 1. British English (left) and Spanish monophthongs (right)[17]

Conceptualising pronunciation: introducing phonemic symbols

Taking the example of English vowels above, if learners do not have any mental reference with which to compare instantiations of these sounds when they hear them, it will be difficult for learners to conceptualise them. By simply offering learners labels with which to categorise the FL phonology, they are implicitly being informed that sounds they consider to be *similar* or *the same* are in fact different. As a case in point, if the only label FL learners have for /a/ sounds is the letter <a> (e.g. for Spaniards), they will associate all *similar* sounds (/æ/, /ʌ/, /ɑː/ and /ə/) to the mental representation they have for their native vowel /a/. If learners realise that what they understood as /a/ can be divided into four sounds, even if they do not perceive differences among them, it is a first step towards subsequent better discrimination and production. By simply providing learners with labels for these sounds, they should be better equipped to categorise them, or as Fraser (2006, p. 20) puts it, 'organise them in their minds'. This would enable learners to start associating future exemplars of these sounds as belonging to those categories, which should in turn help them notice relevant phonetic features in those sounds and consequently strengthen/create phonetic categories for them.

Furthermore, as noted above, one of the most challenging aspects in English pronunciation is the lack of a one-to-one correspondence between phonemes and graphemes. As Roach (2000, p. 1) suggests, 'because of the notoriously confusing nature of English spelling, it is particularly important to learn to think of English pronunciation in terms of phonemes rather than letters of the alphabet'. This is especially convenient for English given that each symbol represents only one sound. Phonetic notation cannot help a person to pronounce the sounds of the foreign language better, but it can show what the

---

[17] The English version has been taken and adapted from Ladefoged and Johnson (2011, p. 90). They represent the acoustic measurements for British English monophthongs from a newscaster. The Spanish version has been taken and adapted from Hualde (2014, p. 122), based on the data from sixteen male speakers by Quilis and Esgueva (1983, p. 244).

proper sequences of sounds are for any given word or sentence of the language (MacCarthy, 1978). Phonetic symbols should help FL learners pronounce words with silent letters or sound-spelling correspondences that do not exist in their native language, such as *centre, Schmidt,* or *Georgia* for example. For example, these words are likely to be mispronounced by Spaniards as *[ˈθentre], *[estʃˈmidt] or *[xeˈorxja]. Nevertheless, by realising that the <tre> in *centre* is pronounced /tə/, or that the <Sch> in *Schmidt* is pronounced /ʃ/, it should be easier for learners to avoid inserting extra sounds.

Learning the symbols makes it possible for people from any language background to understand phonemically transcribed words and to be able to pronounce them even if they have not heard them before. As Mompean writes (2005, p. 1):

> The use of phonetic symbols in foreign language teaching and learning is potentially very advantageous. Provided that the values of phonetic symbols are known and that the foreign language learner can produce and discriminate the sounds symbols stand for, these advantages include, among other things, increased awareness of L2 sound features, 'visualisation' of such intangible entities as sounds, increased learner autonomy when checking pronunciation in dictionaries, etc.

Hence, if students are able to create adequate mental representations for those intangible entities, they will be in a much better position to judge their own performance and act as self-evaluators. It will be easier for them to judge whether what they are producing is similar to what they have in mind or not (i.e. sound categories) and, through guided listening and extensive practice, work towards the eventual mastery of the sound system of English.

This idea of mental category formation is in line with the recommendation of raising learners' awareness of how pronunciation works in order to empower them with self-monitoring skills (see e.g. Couper, 2003). If learners understand how pronunciation works and the targets they are supposed to achieve (i.e. if they have clear mental categories for the target sounds and patterns), they should be in a better position to monitor their pronunciation and continue improving autonomously. This is especially convenient given that even state-of-the-art technology is not yet suitable as to evaluate learners' mistakes automatically (see section 1.3.2b below).

Self-monitoring pronunciation

Numerous authors have emphasised the importance of self-monitoring for pronunciation development (e.g., Acton, 1984, 1997; Avery & Ehrlich, 1992; Dickerson, 1987; Firth, 1987; Lord, 2005; Pennington, 1996; Yule, Hoffman & Damico, 1987).[18] When learners are able to monitor their performance, they can continuously evaluate their progress, which allows them to notice their errors and be aware of those aspects that need further work. As Avery and Ehrlich (1992, p. 219) put it:

> While the individual instructor plays a critical role in the initial stages of pronunciation improvement, it is the individual student who must ultimately take responsibility for ongoing improvement. […] Self-correcting and self-monitoring minimize dependence and maximize self-reliance, allowing students to continue pronunciation improvement outside the classroom.

Dickerson (1994, p. 19) emphasised the importance of teaching learners to predict sound-spelling correspondences by re-writing the saying 'give a man a fish and he will eat for a day, but teach a man to fish and he will eat for a lifetime' as 'teach someone the sounds of a word, and that person can say that word. But teach someone to predict those sounds, and that person can say any word.' The saying refers to one particular aspect of pronunciation, but it could be extrapolated to pronunciation teaching and learning in general.

As recommended above, raising the learners' awareness of the way the FL phonology works should empower them with self-monitoring skills that help them work on their pronunciation autonomously. For example, if Spanish EFL learners have adequate categories (or mental targets) for /s/ and /z/ and they realise that they do not usually pronounce the distinction between the two, they should be able to pay attention to subsequent instantiations of /z/ they hear – in theory, facilitating reinforcement of this perceptual category. This should, in turn, enable them to devote particular attention to the pronunciation of that sound in their output and work towards attainment of the target they have in mind. As a case in point, Derwing et al. (1998) reported that learners receiving segmental training showed low ratings of accentedness when they could monitor their performance, something useful in order to self-correct mispronunciations that may cause communication breakdowns.

---

[18] See also Couper (2003), Derwing and Rossiter (2002), Hardison (2005), Jones (1997), Morley (1991, 1994), Smith and Becknamm (2005, 2010).

Monitoring pronunciation can be considered to be more challenging than monitoring other components of the FL, such as grammar. Monitoring the latter will be possible insofar as learners have some declarative knowledge about how the language works, whereas pronunciation requires learners to be able to perceive the FL adequately before making any judgements. Nonetheless, possessing declarative knowledge about the way pronunciation works is thought to be beneficial in order to foster noticing of relevant features of the FL phonology.

The role of focus on form and explicit instruction in FL pronunciation

Naturalistic approaches to FL teaching and learning have commonly held that foreign languages can be *acquired* in the same way as people acquire their first language, placing a very high value on language exposure and deemphasising the role of explicit instruction (e.g. Krashen, 1979). Nonetheless, in the same way that mere exposure to the language has been found to be insufficient in order to acquire certain grammatical and pragmatic aspects of the language properly (Swain, 1985, 2005), research has shown that foreign or second language learners often retain some degree of foreign accent even after having spent several years in a foreign country (see e.g. Bongaerts et al., 1997; Flege, 1991, 2009; Moyer, 1999). This suggests that mere exposure is not enough for pronunciation acquisition either. In fact, pronunciation has been considered to be 'relatively immune to all but the most intensive formS-focused treatments' (DeKeyser, 1998, p. 43).[19]

Focus on form (Long, 1991) refers to the need to direct learners' attention to formal features of the language that would otherwise go unnoticed. As Tomlin and Villa (1994, p. 184) note:

> Humans […] are bombarded constantly with overwhelming amounts of sensory and cognitive information. It is the human attention systems that reduce and control that influx of information […] attention is employed to help sort out that input and to bring order to the chaos […] overwhelming the learner.

Thus, authors have stressed the role of attention and noticing in FL learning (Guion & Pederson, 2007; Schmidt, 1990). Noticing may refer to learners' realising that their

[19] A distinction should be drawn between focus on form, focus on formS and explicit instruction (Long, 1996). Focus on 'formS' is mostly concerned with linguistic forms in isolation in some kind of structural syllabus, and 'focus on form' has to do with learners' attention being directed to the language as an object, but as part of a bigger context where attention is directed towards both form and meaning. FFI will be used here to refer to instruction that directs the learners' attention to language form, not necessarily in an explicit way, while explicit instruction (EI), as the name indicates, will be used to refer to training including overt explanations about pronunciation features. See Williams (2005) for a revision of the different interpretations of the concept of 'focus on form'.

language is different from the target, that is, noticing a 'gap' between their productions and a target model (Swain, 1985), or to the detection of features previously overlooked in the input (Schmidt, 1990, 2010). The former case refers to the learners' need to realise gaps in their knowledge in order to be able to 'fill' them. The latter refers to the need of gaining awareness of a particular target feature in order to assimilate it. In Schmidt's (2010, p. 722) words, 'input does not become intake for language learning unless it is noticed, that is, consciously registered'.

This is especially important for pronunciation, as learners are commonly unable to perceive the FL correctly, which is aggravated by the fact that it is difficult to process form and meaning simultaneously (see Guion & Pederson, 2007). There is growing consensus that instruction is most effective when it includes attention to both meaning and form, especially given the limitations in classroom time. As Thomson (2011) points out, because learners' cognitive resources are limited '[i]f learner attention is oriented toward phonetic information, more of the input can be incorporated into emerging L2 categories' (p. 746).

In this regard, explicit instruction provides a suitable and efficient way of raising learners' awareness of features in the input that would otherwise go unnoticed. Researchers have long advocated the role of explicit instruction in SLA (Ellis, 1993; Gass, 1997; DeKeyser, 2003 – see Doughty & Williams, 1998 for a review), as overtly directing the learners' attention to the formal features of the language allows for a 'more efficient use of [the] limited exposure to the sounds, words, and sentences of the language [they] are learning' (Spada & Ligthbown, 2008, p. 182).

Ellis (1997) illustrates the importance of explicit knowledge with the use of the subjunctive in French, arguing that someone who knows about the subjunctive is better equipped to notice it in the input (s)he receives. Following Ellis's example, by knowing about the subjunctive in French, learners will be able to observe how native speakers use it in context, reinforce their knowledge about it by observation and then start implementing it in their own speech. This becomes all the more relevant for pronunciation, given that by being aware of segmental or suprasegmental aspects of the FL phonology, students will be in a better position to identify these aspects every time they are exposed to language and subsequently start using the features themselves. Using the example above with the problems FL learners face with English vowels, if learners do not know about /æ/, /ɑː/ and /ʌ/, they cannot notice them when they hear instantiations of these sounds in native speech; they will identify all these sounds as belonging to the only sound they are familiar with, namely /a/. However, by being aware of the existence of the sounds, learners are better prepared to notice them when they hear them; and consequently, they can associate every instantiation they hear of a sound to its corresponding 'label' or category.

The fact that form-focused training on specific pronunciation features can bring about benefits is not in question any longer (see section 1.3 below). However, few studies have isolated the impact of explicit information on pronunciation training. In this regard, the studies by Saito (2013) and Saito and Lyster (2012) shed light on the positive effects of explicit instruction as compared to focus on form alone. While Saito and Lyster (2012) found beneficial effects of form-focused instruction (FFI) when it was accompanied with corrective feedback from the teacher, receiving FFI alone did not yield significant gains as compared to a control group. Furthermore, when a group receiving explicit instruction came into the equation (Saito, 2013), this was found to be significantly more advantageous, as the improvement in the FFI group was only moderate and context-dependent, whereas the group receiving FFI and explicit instruction obtained larger pronunciation gains and generalised to new lexical contexts. Hence, Saito concludes that FFI alone does not seem to be enough to trigger learners' awarenes of phonetic information, as learners were not able to show improvements in items that had not been shown in training. As Saito (2013, p. 4) recommends:

> In conjunction with the default L2 speech learning process, pronunciation instruction should […] be designed to (a) raise learners' perceptual noticing and awareness of new sounds not only at a lexical level but also at a phonetic level to promote the formation of new phonetic categories and (b) encourage them to practice the new L2 sounds in output to enhance their production abilities at various processing levels in relation to the present state of their developing mental representation.

Starting from the assumption that adequate perceptual categories guide subsequent productions of the FL phonology (Flege, 1991, 1995), and after numerous studies showing that training perception also improves production (e.g. Bradlow et al., 1997; Lambacher, et al., 2005; Motohashi-Saigo & Hardison, 2009; Thomson, 2011; Wang & Munro, 2004), it can be hypothesised that improvements in the learners' perceptual domains should help them monitor their productions (Thomson, 2011). As mentioned above, by having clear models against which to compare their productions, learners will be in a better position to judge their own pronunciation and devote special attention to areas that they feel need to improve. In fact, the learners' ability to perceive their own speech accurately has been found to play a very important role in their accurate production of the FL phonology (Baker & Trofimovich, 2006).

1.2 CALL: Computer-Assisted Language Learning

Language learning has long been assisted by computers. As a field, CALL is incredibly interdisciplinary, benefitting from various other fields, such as Computational Linguistics, Human-Computer Interaction, Applied Linguistics, Artificial Intelligence, Instructional Technology and Design, or Psychology (Levy, 1997a). The primary purpose of CALL has always been to facilitate or enhance language learning thanks to the possibilities offered by computers. As Levy and Stockwell (2006, p. 2) observe,

> historically, the invention of new technologies has been largely motivated by a desire to extend or to overcome our innate limitations as human beings, especially those set by our physical or mental capabilities. Thus, new technologies such as the pen and paper for writing aid our memory, the telescope or microscope enhance our vision, the telephone extends our ability to communicate at a distance, the hammer amplifies our strength, and the car or airplane extend our range.

Because the focus of this dissertation has to do with the teaching and learning of pronunciation with technology, it could be considered to fit best within CAPT (Computer-Assisted Pronunciation Training), as CALL refers to language learning in general. However, since both fields are grounded on the affordances offered by technology, they share a number of common features. Consequently, although a brief summary of the general advantages of CALL will be presented here, other basic advantages of technology, such as limitless practice or immediate feedback will be discussed in section 1.3 (CAPT) to avoid repetitions.

Saving time is one of the major advantages offered by technology. This can be appreciated in routinely tasks such as designing lessons and teaching. Sharma and Barrett (2007) mention interactive whiteboards that allow the teacher to save a lesson and then recycle it with another group, saving preparation time for the second lesson. Simply using PowerPoint (or a similar software) can save hours and hours of work (and classroom time), freeing the teacher from writing explanations and examples on the board. By being able to present all the information at once, teachers can use the time formerly devoted to writing explanations and examples to offer examples or practice. Moreover, explanations and examples can be greatly enhanced by supplementing them with pictures and other multimedia.

Thanks to the variety of presentation styles and interactive designs that incorporate multimedia, current software holds the potential to be highly motivating, one of the main factors likely to influence foreign language learning in general and mastery of pronunciation in particular (Elliott, 1995; Moyer, 1999). Furthermore, with current technologies like mobile apps, for example, this potential for motivation can be further exploited with the appropriate gamification[20] techniques (see Zichermann & Cunningham, 2011). These may range from implementing a scoring system with the appropriate rewards (be it trophies, extra levels, online 'prestige', etc.), to using a leaderboard showing the latests rankings, or even sharing the scores obtained through social social networking services (SNSs) such as Facebook or Twitter.

Focusing on FL learning, one of the clearest advantages of computers is that they offer learners the possibility to contact and communicate with native speakers, being exposed to authentic language that would otherwise be missing in the classroom. Something some have defined as 'an economical and convenient alternative to travel' (Hoffman, 1996, p. 68). Thanks to Computer Mediated Communication (CMC), today's learners can have tea with a friend abroad without leaving their homes, therefore being able to interact with native speakers without spending a penny. They can practise their written skills with native speakers by chatting through SNSs, forums, messaging services (such as the old MSN, Whatsapp, or Skype – most of which can also be accessed from smartphones), or they can also have a synchronous (video) talk with someone miles away thanks to free voice over internet protocol (VOIP) services such as Skype or FaceTime. All this brings incredible advantages from tandem learning (see Horgues & Scheuer, 2015) to the learners' fingertips. In fact, online language learning platforms already exist in which learners act as both students and evaluators; students are given points for acting as native-speaker evaluators of students learning their L1 and then these points can be redeemed by corrections by other native-speakers (e.g. LiveMocha).

Someone who wanted to learn a language a few decades ago had to either go to the country where the language was spoken or buy videotapes or cassettes in order to obtain samples of language. However, both options entailed a considerable investment on the part of the student and they were not immediately available. Today, learners have instant access to limitless samples of language that are freely available on the internet (podcasts, videos on

[20] Gamification refers to the idea of applying typical elements of game playing and game mechanics (e.g. leader boards, scoring systems, prizes, score sharing, etc.) to other activities in order to engage audiences.

YouTube, songs, online radio, films, etc.), not only offering access to a wide range of topics, but also to a wide range of voices and accents (something particularly promising for pronunciation instruction). Moreover, thanks to today's smartphones this can be done anywhere, at any time. Today's FL learners can look up the meaning of words from their phones with dictionary apps, check how certain structures are used in context using any search engine, or even find pictures that exemplify words they need to look up (e.g. by typing 'hummingbird' on Google and getting hundreds of images illustrating what a hummingbird is).

A further argument in favour of adopting technology for teaching is the fact that today's learners are considered to be 'digital natives' (Prensky, 2001). That is, they have grown up surrounded by technology and they are used to using technology for everything. Although it has yet to be tested empirically, Prensky (2001) maintains that today's students even process information differently as a result of their experience with technology. In Prensky's description of today's learners, graphics are preferred over text, they like to multi-task and they are used to accessing information really fast. He claims that if these learners can memorise all the Pokémon characters (more than 100) with all their characteristics, there is no excuse not to learn other things such as populations, capitals and nations in the world. His argument is that it all depends on the way information is presented, which clearly has an impact on the learners' motivation. Hence, exploiting technologies in order to offer information will hopefully help engage and motivate digital native learners, who may feel more attracted to practise the past tense with an online quiz rather than with a paper-based activity, for example. A great example of innovation is the Digital Kitchen project developed at Newcastle University (Seedhouse et al., 2013, 2014), where students learn to cook a number of recipes at the same time that they learn to communicate in the foreign language through task-based learning.[21]

Finally, something particularly interesting for research is the fact that in CALL and CAPT the computer acts as both a tool for instruction and a tool for research, allowing teachers and researchers to collect comprehensive and detailed data of learners' performance (Chun, 1998). This has obvious advantages in order to monitor students'

[21] The kitchen incorporates a number of devices attached to every appliance and instructions only proceed when the learner takes the right appliance and performs the action (s)he is supposed to perform. This holds great potential for learning vocabulary in quite an authentic manner though task-based learning. It was originally designed to learn French but it is being adapted to other languages (see http://digitalinstitute.ncl.ac.uk/ilablearnkitchen/).

weaknesses and progress over time, being able to collect information from individualised performance even when handling big groups of students.

This section has briefly discussed some of the advantages offered by computers for teaching and learning in general, and for language learning in particular. In the next section the focus is on the specific benefits of CAPT. As noted above, given the various challenges foreign language learners face when approaching pronunciation, the possibilities offered by technology seem especially suitable for enhancing pronunciation instruction. Hence, section 1.3 presents some of the benefits of CAPT reported in the literature, discusses some of the limitations of current technology for pronunciation training and offers recommendations for future research.[22]

1.3 CAPT: Computer-Assisted Pronunciation Training

As Llisterri (2007) points out, because CAPT requires quite specific resources that were not widely available in conventional computers not so long ago, and given the general neglect pronunciation has suffered in foreign language teaching, CAPT is comparatively much more recent than CALL. However, over the last decades numerous researchers have explored the ways in which pronunciation training can be enhanced through technology.

1.3.1 Advantages of CAPT

One of the major obstacles people encounter when learning a foreign language is the scarcity of input available, as exposure to the foreign language is often restricted to classroom time. However, this traditional problem is mititaged by resources such as digital television or the Internet, allowing today's learners to watch films in the original language, watch interviews or videoclips on YouTube, or download a virtually unlimited amount of podcasts (featuring different accents) from their mobile phones. Nonetheless, technology does not only make materials more *physically* accessible, but also more *cognitively* and *psychologically* accessible (Pennington, 1996).

One of the most significant advantages offered by technology is that it holds the potential to enhance learning in ways that were unthinkable before, something extremely advantageous for pronunciation given the learners' perceptual constraints as a result of their L1. For example, a 3D representation of the mouth illustrating sound articulation can

[22] An abridged version of section 1.3 has been published in Fouz-González (2015), a book chapter dealing with trends and directions in Computer-Assissted Pronunciation Training.

be very helpful for learners who are not able to perceive the correct place of articulation from the auditory signal alone (e.g. English /t/ as alveolar, rather than dental, as in Spanish).

Furthermore, technology makes it possible for students to have individualised and limitless autonomous practice. One common claim why pronunciation instruction does not always find its place in the FL classroom is due to time constraints. As Levis (2007) points out, different learner needs can make group instruction irrelevant for some, particularly in classes with learners with different L1s. However, even though groups of students with a shared L1 may benefit from practice addressing common L1-related problems, each learner will differ in terms of the areas they need to reinforce (O'Brien, 2006). For example, if a student does not perceive the contrast between two sounds (e.g. /æ/ and /ʌ/), the teacher may exemplify it many times and try to offer as much practice as possible, but the class will get to an end and the teacher will get exhausted; besides, not every student will have the chance to practice what they need, as there is only one teacher for a group of students with different needs. Technology, on the other hand, allows students to repeat problematic contrasts as many times as they need.[23] Moreover, provided that the adequate software is available, they may even hear the contrast pronounced by a number of different voices, and in different accents and contexts, something which is not as easy for a teacher to provide and which is especially important for pronunciation.

Another advantageous aspect of CALL and CAPT is the possibility to offer immediate, automatic and tireless feedback on the learners' performance. When evaluating other aspects of language learning (e.g. grammar or lexis), feedback can be provided in the form of a tick or a cross indicating whether learners have got their answers right or wrong, hopefully offering the solution if their answer was wrong. In the case of pronunciation, the same type of feedback can be offered on the learners' perceptual abilities according to the choices they make, but automatic evaluations of the learners' oral production can also be obtained thanks to Automatic Speech Recognition (ASR) technology (see section 1.3.2b). Even though the feedback learners can receive through ASR technology is still far from perfect (or sometimes even adequate, as we will comment below), the fact that every learner can receive individualised comments or advice on their performance is something that a teacher cannot always do, partly as a result of the time constraints mentioned above,

---

[23] As a case in point, Probst, Ke, and Eskenazi (2002) mention that one of their participants repeated the same sentence 61 times. Even though teachers may find some time to offer individual practice, it will never be that much.

but also because not every teacher will be in a position to offer accurate judgements on learners' mispronunciations. The idea of learners receiving feedback on their performance whenever they want and wherever they need is extremely auspicious, especially given the importance of instant corrective feedback in order to avoid fossilisation.

Additionally, the fact that students use technology to practise autonomously with a computer, tablet, or smartphone is also positive for teachers. By keeping a record of learners' performance, the teacher will be able to see how students progress on the different activities proposed and analyse which areas require further attention (Neri, Cucchiarini, Strik, & Boves, 2002). Also, with the appropriate artificial intelligence, extra activities could be selected automatically by the machine based on the learners' performance, offering extra practice on those aspects that are most problematic for a particular learner. This should be easily done with an algorithm that detects students' wrong choices in tasks where they have to choose, but with advanced ASR technology, they could even point to problematic areas in terms of production (e.g. recurrent mispronunciations of a phoneme, unnatural pausing or rhythmic patterns, etc.).

Besides, given that pronunciation is a psycho-motoric action that implies coordination and control over many muscles, the private environment that technologies offer seems an ideal solution for students who need a great deal of repetition and are afraid to do so in front of other people (Witt & Young, 1997). As Chapelle and Jamieson (2008) write, 'what once involved embarrassing repetitions in whole-class oral exercises can now be conducted by learners working individually at the computer' (p. 169). One of the aforementioned constraints on pronunciation acquisition was shyness, as many students avoid speaking in class because they do not dare to do so in front of their classmates (Morley, 1994). Thus, technology offers the private environment learners need to practise comfortably, at their own pace and wherever they feel relaxed. This should allow students to build up confidence by practising on their own and be better prepared to face interactions in a classroom or real situation, thus helping learners overcome foreign language anxiety (Murray, 1999; Pennington, 1999).[24] In Pennington's (1996, p. 8) words, 'computer training systems for pronunciation seem to be one of the few environments where learners' consciousness of their own spoken language errors does not interfere with their learning'. This does not mean though, that classroom practice should be eradicated.

[24] See e.g. the learners' positive comments about their increased self-confidence after Hardison's (2005) training.

Learners could work individually on the aspects they find most problematic in order to increase their self-confidence and then move progressively to more extemporaneous practice, be it in the classroom or in authentic conversations outside the classroom.

In the same way it is commonly argued that teachers should never be replaced by technology, some also argue that CAPT should become indispensable and never be replaced by the teacher (Thomson, 2011). Given the potential advantages of CAPT, technology can enhance the teaching of certain features in ways that traditional classrooms cannot; for example, by training the learners' perceptual abilities with numerous instantiations of a particular key contrast in different positions and by different talkers, or by illustrating sound articulation with talking faces.

1.3.2 Enhancement of pronunciation training through CAPT

The question remains as to how technology can enhance the teaching and learning of FL pronunciation. Apart from the aforementioned advantages of CALL, such as individualised and limitless practice, ease of accessibility, or the provision of instantaneous feedback, there are numerous benefits technology offers for pronunciation training.

As mentioned above, the acquisition of pronunciation involves perceptual and productive abilities. FL learners have to first be able to perceive the FL phonology adequately; second, be able to perform the necessary articulatory movements to attain the FL sounds and patterns in production; and third, manage to produce these sounds fluently, naturally and effortlessly. The sections that follow are arranged according to the ways in which pronunciation training can be assisted by technology, starting with perceptual enhancement, followed by productive enhancement, and finally, discussing some of the limitations of current technology and offering directions for future research. Subsections will cover some of the possibilities and findings obtained from studies investigating each technology, together with some limitations and directions for future research. Because the aim of this section is to review how technology can facilitate pronunciation training, specific courseware or applications will not be addressed here, only the underlying technology.

a) Perceptual Enhancement

The ways in which technology can facilitate the learners' perception of the FL phonology are varied. Perception can be enhanced by simply providing learners with adequate pronunciation models, by offering them the input they need, often unavailable in FL contexts. Furthermore, given the learners' inability to perceive the FL phonology adequately as a result of their L1 interference, perception can also be facilitated by by making features in the input more salient so that learners notice them. Auditorily, this can be achieved by training learners with highly variable stimuli or by using synthesised or modified natural speech, for example. Additionally, perception can also be augmented visually, either by using speech visualisation technology to represent the speech signal or by showing learners articulatory information so that they can perceive sounds and patterns better.

Synthetic speech

One way of making pronunciation more physically accessible is by using artificial (or synthesised) speech. There are two main types of speech synthesis: text-to-speech and concatenative. The former creates artificial speech from scratch and the latter employs recordings of real speech that are edited and cut into small parts and then recombined to form words and sentences.[25]

Artificially generated speech offers promising prospects for FL learning, as it allows learners to find out how *any* word is pronounced by simply typing it on a computer, phone or similar device without needing natives speakers to pronounce it. The possibility of listening to the pronunciation of a word is something every electronic dictionary offers. However, the advantage with speech synthesis is that words do not need to be pre-recorded in the system.[26] Additionally, while audio illustrations of pronunciation in most learner dictionaries are restricted to words in their citation form, synthetic speech can exemplify the pronunciation of past tenses, plurals, or even proper nouns.[27] Hence, this is not only advantageous for language learners, who can find out about the pronunciation of a word by simply typing it, but also for materials developers, who do not need to resort to native speaker models.

Research has often focused on the comprehensibility of synthetic speech as compared to natural speech, as well as on its potential applications. This technology has drawn many researchers' attention given the possibilities this technology offers for

---

[25] For a detailed account of how speech synthesis works see Taylor (2009).

[26] As compared to other courseware that requires the authors to pre-record every word or utterance that want to include in the package, speech synthesis allows for the creation of an unlimited number of utterances.

[27] Although pronunciation dictionaries do include written illustrations of how plurals, past tenses or weak forms should be pronounced, audio illustrations of such aspects are often missing (see e.g. Jones, Roach, Setter, & Esling; Wells, 2008).

everyday tasks such as GPS navigation, automatic answering machines or applications that convert text into speech for the visually impaired, among others (see Winters & Pisoni, 2004 for a review).

As far as CALL is concerned, Handley and Hamel (2005) identified three types of text-to-speech synthesis applications, namely reading machines, pronunciation models, and conversational partners. The first two functions are rather similar in that both would imply a transformation from text into speech that would somehow serve as a model to the learner (like audio illustrations of pronunciation available in dictionaries, or activities in which a text or dialogue is read by the machine so that students can listen to its pronunciation). An example of the third function could be the use of synthetic speech in talking heads or virtual worlds, where learners can interact with animated agents in simulated conversations with the machine. As Handley (2009) notes, this technology allows computers to generate an unlimited number of responses in dialogues that may be unpredictable, something that would be much more difficult and costly with recordings of natural speech. The problem is, though, that for a machine to be able to converse with a human, it requires much more than being able to produce utterances. On the one hand, the machine needs to 'understand' what users say by means of ASR technology (see below), and on the other hand, it should be able to create responses for *any* question that users may ask – which would require pre-planned responses or very advanced artificial intelligence capable of having a spontaneous conversation with a human.

Motivated by the importance of listening skills in language learning, Kang and his colleagues recommend synthetic speech as a means of palliating the lack of input some learners experience, as this would allow teachers to create their own listening materials in the absence of authentic input (Kang, Kashiwagi, Treviranus, & Kaburagi, 2008). Although they maintain that today's learners have little opportunity to receive input, something somewhat surprising given the vast range of free podcasts available worldwide, the fact that synthetic speech allows students to hear the pronunciation of *any* word out of text is an added value to many CALL applications. Kang et al. (2008) found that listeners' evaluations of synthetic voices were strongly correlated with the extent to which they understood the audios; that is, voices were rated positively as long as they were understood. However, their data do not offer an answer to the question of whether synthetic speech is as effective as natural speech in training listening skills. They found that synthetic speech obtained even better comprehension scores than natural speech at the word level. Nevertheless, whether learners could improve their listening skills by means of artificial speech remains unanswered, as no pre/post-test evaluation was carried out in order to measure the learners' abilities.

In a similar line, Hincks (2002) has argued that synthetic speech might be suitable in light of proposals of phonological models of English for international speakers, such as Jenkins' (2000) Lingua Franca Core (LFC), given that there do not exist native speakers of this variety and synthesisers could act as models. The LFC has been proposed as a selection – or prioritisation – of those elements that are supposedly essential for intelligibility in the most common varieties of English. Hence, it recommends avoiding the teaching of features that may be difficult for foreign language learners and do not affect the communicability of their messages. Nonetheless, in spite of the potential of synthetic speech to illustrate stress placement in a word, or how a word is pronounced (in terms of *what sounds* a word is composed of – things that do not demand very high quality synthesis), research is needed to shed light on whether someone could acquire a new phonological system based on artificial speech.

In spite of the above-mentioned possibilities speech synthesis offers, for an effective implementation of artificial speech as pronunciation model, there are several aspects research needs to address. Something worth considering is whether learners would be able to identify in real speech the aspects they practise with artificial speech. For example, connected speech processes such as linking, coalescence, or merging; suprasegmental phenomena such as contrastive intonation or a 'natural' use of rhythm; or other aspects of real speech that affect segmental features, such as elisions or assimilations. Even though being comprehensible enough may be an acceptable standard for devices that use synthetic speech for tasks such as giving instructions or directions, standards should be set quite high when using this technology for language teaching, and especially for pronunciation teaching.

As pointed out by Schwab et al. (1985), even high-quality synthetic speech is not as easy to perceive and understand as natural speech (e.g. Pisoni, 1982; Reynolds & Jefferson, 1999). However, since research has shown that listeners can be trained to understand synthetic speech better (Schwab et al., 1985), Schwab and colleagues recommended training listeners to understand artificial speech in the absence of high quality synthetic speech. The problem is that artificial speech can be more cognitively demanding than natural speech, as research suggests that it usually requires more processing time (Pisoni, 1981; Reynolds &

Jefferson, 1999). [28] Furthermore, in the context of foreign language pronunciation acquisition, merely being able to understand the model is far from enough. Understanding your instructor would be the first, obvious, and essential step; the question is whether an artificial 'instructor' could serve as an adequate model and help learners create more native-like (human) categories. Perhaps speech synthesis comes in handy to illustrate that the <ch> in *archives* is pronounced /k/ and not /tʃ/ (a common mistake Spanish learners make – see chapter VI), or even the fact that *cat* is pronounced /kæt/ and not /kʌt/, but this does not mean that by being exposed to this type of input learners are able to create new phonetic categories for these sounds, facilitating either perception or production.

As regards segmental acquisition, new phonetic categories are thought to develop after exposure to numerous instantiations of the sound in question (Flege, 1991, 1995). In this respect, extensive auditory training with exposure to natural variability has proven useful for improving perception of FL key contrasts (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993), improvements that encourage long-term modification of learners' perception even in FL contexts where learners are not often exposed to the contrast (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994) and which can be transferred to production (Bradlow et al., 1997). Nevertheless, whether listening to an artificial, sometimes robotic voice can facilitate the acquisition of natural speech in humans is subject to empirical validation. Indeed, research suggests that imitating a voice that is similar to one's own is better than imitating a dissimilar voice (Probst et al., 2002). Therefore, perhaps a synthetic voice completely generated from scratch or concatenated, may not be a good match for a human.[29]

The results of the first attempts to train listeners to perceive difficult sounds with synthetic stimuli were not as encouraging as expected. Despite the fact that gains were obtained after training with synthetic speech, these did not transfer to perception of natural speech (Strange & Dittman, 1984). Nevertheless, subsequent studies have found positive correlations between the perception of synthetic and natural speech (e.g. Pruitt, Kawahara, Akahane-Yamada, & Kubo, 1998; Yamada & Tohkura, 1992), with gains that can be transferred to novel stimuli and retained over time (Wang & Munro, 2004). As noted above, synthetic speech has also been implemented in talking faces that illustrate

---

[28] For a comprehensive review of research on the perception and comprehension of synthetic speech see Winters and Pisoni (2004).

[29] Synthetic voices have been described by some students as sounding 'robot-like' or 'artificial' (see Massaro, Cohen, Tabain, Beskow, & Clark, 2012).

articulatory information of FL sounds (Chen & Massaro, 2011; Massaro & Light, 2003; Psyentific Mind, 2012) or act as conversational partners (Wik & Hjalmarsson, 2009). However, very few studies explore whether learners actually improve their pronunciation after using this software (see e.g. Massaro & Light, 2003), the majority focus on comparisons of comprehensibility ratings between natural and synthetic speech.

The fact that a model speaker is not needed in order to record all possible combinations of sounds to form words makes it easier for non-native teachers to create materials, allowing them to save money, time and effort. However, future research should address empirically whether synthetic speech can be effective for the acquisition of FL phonology. Investigations should not be limited to shed light on whether using synthetic voices for extensive training results in better discrimination or identification of the sounds in question; they should also address whether listening to a synthetic voice for extended periods of time (as we do when we listen to native speakers in real life) also helps to acquire new phonetic categories for the FL phonology – given that one of the uses advocated by researchers is working on listening skills (Kang et al., 2008).

Additionally, future investigations could also establish comparisons between completely synthetic and concatenated speech. While the former may be suitable to illustrate the phonological composition of words (Chun, 2007), or lexical stress (Hincks, 2002), the latter may be more appropriate for phonological acquisition, as it resembles natural speech more given that it is obtained from natural speech. Finally, research should also address the potential of this technology for teaching suprasegmental aspects of speech. To date, even top-notch synthesis sounds rather artificial in terms of rhythm or intonation (e.g. Siri, iPhone's virtual assistant). Research suggests that synthetic speech does work for teaching word stress (Hincks, 2002), but it would be interesting to explore whether it could be used as a model for other aspects of prosody, such as rhythm, attitudinal or contrastive intonation. Research should assess current state-of-the-art synthetic speech in order to test its potential to convey suprasegmental information.

Furthermore, an evaluation of speech synthesis systems is paramount before implementing the technology on CALL applications, given that general purpose synthesis systems are sometimes implemented in applications that need much higher standards, such as pronunciation models or conversational partners (Handley, 2009). There are some flagrant examples of speech synthesis models where it seems as though designers had not even tested the product before releasing it. As a case in point, one of the few smartphone apps which are specifically devised for pronunciation instruction offers synthetic representations such as */aɪ el el gəʊ/ for the sentence *I'll go*; */aɪ em ˈbɪzi/ for *I'm busy*; or /juː el el rɪˈgret ɪt ɪf juː ˈdəʊnt tiː gəʊ/ for *You'll regret it if you don't go* (Fouz-González, 2015). Surprisingly, this app does not perform so poorly because it allows learners to type improvised utterances; these are model sentences offered by the app that had been deliberately included in the design.

Auditory enhancement

The previous section explored the possibilities offered by synthetic speech as a source of input, the first aspect that learners need in order to improve their perception of the FL phonology (i.e. enhancing *physical accessibility*). This section concentrates on the ways in which technology can enhance the input learners receive in order to make certain features more salient. Given that one of the major obstacles in FL phonological acquisition stems from the learners' perceptual bias as a result of their L1 experience (see section 1.1.2), one of the advantages technology offers is to facilitate the learners' perception of aspects that often go unnoticed.

A traditional technique to train learners' perception of problematic FL contrast is the High Variability Phonetic Training (HVPT) paradigm put forward by Logan et al. (1991). It consists in exposing learners to multiple exemplars of the target sounds as pronounced by different talkers in different phonetic contexts. This variability is supposed to offer learners numerous instantiations of the 'target category' learners need to acquire, one of the pre-requisites for the acquisition of FL phonemes (Flege, 1991, 1995). The technique has been tested numerous times with positive results (e.g., Carlet & Cebrián, 2014; Iverson, Hazan & Banister, 2005; Logan et al., 1991; Lively et al., 1993; Rato, 2014), and gains in perception have been shown to transfer to production (e.g., Bradlow et al., 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Motohashi-Saigo & Hardison, 2009; Thomson, 2011; Wong, 2015).

But, in what other ways can technology enhance the learners' perception of the FL phonology? An interesting approach is the acoustic enhancement of important elements in the speech signal in order to make relevant features more noticeable. That is, if learners cannot 'hear' a contrast, the latter is made more salient by enhancing the acoustic signal. There are several ways in which speech can be enhanced acoustically. Researchers have explored the potential of artificially generated or synthesised speech (Hincks, 2002; Jamieson & Morosan, 1986; Strange & Dittmann, 1984), combinations of natural and synthetic speech (Wang & Munro, 2004), as well as acoustically enhanced natural speech (Barreiro-Bilbao, 2013; Hazan & Simpson, 1998, 2000).

Since listeners pay attention to the same acoustic cues when listening to natural and synthetic speech (Gordon, Keyes, & Yung, 2001), the synthetic enhancement of acoustic cues that are difficult to perceive for FL learners may help them perceive these cues more easily and improvements should also transfer to natural speech. As a case in point, Wang and Munro (2004) found that FL learners can improve their perception of non-native vowels (namely /i/-/ɪ/, /u/-/ʊ/, /e/-/æ/) by training with a combination of synthetic and natural stimuli. They employed Jamieson and Morosan's (1986) 'fading technique', exposing learners to synthesised and natural examples of the same word with modified vowels of varying lengths in order to draw their attention away from vowel length and help them focus on vowel quality instead. They used synthetic and natural stimuli produced by different talkers in order to help learners gradually distinguish among these categories. Participants improved considerably on the three pairs of sounds (14 points for /i/-/ɪ/, 32 for /u/-/ʊ/, and 16 for /e/-/æ/) and their ability to identify these contrasts correctly was generalised to novel stimuli and retained three months after training.

With a different goal in mind, Hincks (2002) used WaveSurfer to help students correct wrong stress placement of two problematic English words by manipulating synthetic speech. Her students worked with synthesised words and the visual aid of a talking head and were told to manipulate them in order to gain awareness of correct stress placement. Although the results were positive, she acknowledges that the method is extremely time-consuming for teaching the pronunciation of isolated words. In fact, research suggests that students are able to perceive and correct lexical stress errors in problematic words with brief, explicit explanations that help them 'notice' the key features (see Fouz-González & Mompeán, 2012; Mompeán & Fouz-González, in press; chapter VI in this dissertation).

Hazan and Simpson (1998, 2000) investigated the effects of speech modification for augmenting intelligibility in noise conditions. Their approach consisted in enhancing the saliency of particular characteristics of the signal that are important for perception by speakers under certain conditions; in their study, the acoustic cues that contribute to consonant identity. They followed Stevens's (1985) concept of 'landmarks', that is, parts of an utterance from which listeners extract information about the underlying distinctive features. Hazan and Simpson claim that, since only a small proportion of the acoustic signal is altered, speech retains a high degree of naturalness while allowing them to increase intelligibility significantly. Even though their method focuses on increased intelligibility in noise, the technique can also be applied to foreing language learning, as augmenting certain characteristics of the speech signal may help learners perceive problematic contrasts. In their study of 2000, Hazan and Simpson tested this technique with Spanish and Japanese speakers and found that, despite individual differences across subjects, the enhanced condition was always perceived better that the natural condition. Moreover, they highlight the fact that their enhancement technique yielded improved intelligibility even though listeners did not receive training, and despite having based the enhancement on their knowledge about cues that affect L1 listeners. They claim that by carefully examining perceptual cues that affect specific L1 populations, the results could be even better.

However, the results of recent studies are not so auspicious. Barreiro-Bilbao (2013) found no additional benefits of enhancing speech for the contrasts /s-z/ and /s-ʃ/ for Spanish speakers. Despite the fact that learners were able to improve under the enhanced condition, the results were generally better when training with natural speech. In a similar line, Akahane-Yamada, Adachi, Kahawara, Pruitt, and McDermott (1998) modified duration cues in natural speech in order to enhance the contrast /r/-/l/ using the STRAIGHT software. They found that scores improved significantly after training both in terms of perception (rising from 44.1% to 97.1%) and production (with an increase from 16.2% to 71.3%), although it is important to note that training was not only auditory, it also involved visual feedback on the learners' productions as well as a visual representation of the model's speech.

Pruitt et al. (1998) compared several types of enhancement for the perception of /r/ and /l/ by Japanese speakers. These included slow and exaggerated pronunciations by human speakers, removing the surrounding phonemic context (i.e. context truncation), or manipulating the spectral and temporal characteristics of the acoustic signal in natural speech. They found that when speakers modified their articulation trying to enhance intelligibility by either slowing it or exaggerating it, identification performance decreased as compared to normal production. The reason they give is that these manipulations may modify natural cues used by NNS to perceive the contrasts. As regards acoustic manipulations of natural speech, they found context truncation to be effective to some extent, although their performance decreased dramatically when reducing the phonetic context too much. Temporal lengthening and spectral enhancement also yielded positive results in one of their experiments, but they point to the need for further research exploring different ways of manipulating the speech signal. In any case, the authors concluded that training with modified natural speech was effective, as it produced learning and it lead to generalisations in non-trained stimuli in natural speech.

In a similar line, Iverson et al. (2005) compared several training techniques for improving identification of /r/-/l/ by Japanese speakers, including HVPT and different types of modified stimuli. They found no significant differences as regards improvements in perception across methods. In light of these findings, the authors argued that there should be nothing particularly special about using natural variability, although they do acknowledge that natural speech is much less laborious and thus probably more convenient, as no particular gains are obtained by using altered stimuli.

The results summarised above offer interesting prospects for CAPT, as one of the major obstacles learners face is the adequate perception of the FL and it seems that training with acoustically modified speech can help FL learners improve their perception of problematic sound contrasts, or at least some of them. However, as in other areas of pronunciation acquisition research, a disproportionate number of studies have concentrated on the English contrast between /r/ and /l/ for Japanese speakers. Research should address other key contrasts and also include other language groups. In this sense, some of the studies cited here offer interesting insights concerning other problematic sounds as well as different L1 backgrounds (e.g. Barreiro-Biblao, 2013; Hazan & Simpson, 2000; Wang & Munro, 2004).

Additionally, as noted in the previous section dealing with speech synthesis, future research should not be limited to investigating whether learners can improve their perception of a problematic contrast after training with modified speech, whether this type of input is adequate for the creation of new (human)[30] categories of the FL phonology should also be addressed. Furthermore, future studies should also investigate whether enhancing relevant cues for suprasegmental features can also result in an improved perception by FL learners (e.g. rhythm or contrastive intonation); to the researcher's knowledge, no studies have been carried out in order to test possible benefits of the acoustic enhancement of suprasegmentals for FL learners. Finally, more research is needed in order to determine whether training perception with acoustically modified stimuli can bring about benefits in production too; the study by Akahane-Yamada et al. (1998) offers

[30] The reference to 'human' refers to whether training with an artificial and sometimes 'robotic' voice can result in the acquisition of natural pronunciation that sounds like that of a native speaker when learners speak in the foreing language – given that they would need to transfer improvements from perceptual practice to the productive domain.

positive evidence in this regard, but given the additional visual feedback offered, it is not possible to conclude whether gains could be obtained from the auditory training alone.

Visual enhancement

Speech perception can also be enhanced visually, for example, by offering learners articulatory information that helps them perceive subtleties in the speech signal; for example, the fact that the English /t/ has an alveolar articulation, rather than dental, as in Spanish, or by facilitating perception of features that are otherwise extremely difficult to perceive through visualisation of the speech signal, such as pitch contours illustrating intonation.

Even though these techniques facilitate the learners' perception of features of the FL phonology, they are often used as a means of providing feedback on the learners' performance, rather than as perceptual enhancement alone. Hence, although they can be considered to be a type of *productive enhancement* (in that they guide the learners' subsequent productions or prompt them to attempt more repetitions), they are included here as types of *perceptual enhancement*, as it is this prior perceptual facilitation (through the noticing of problematic features of their speech in the feedback received) that boosts potential gains in production.

Visual enhancement of acoustic cues: visual speech

An extensively explored area of research in CAPT is speech visualisation, that is, the visual representation of the acoustic properties of speech. The techniques explored in this section facilitate the perception and interpretation of acoustic information (sometimes perceived incorrectly by non-native speakers due to the aforementioned factors) with support from a channel different from the auditory.

As pointed out above, they are often used in order to facilitate the learners' perception of their mispronunciations, rather than to enhance the learner's perception of a model's speech. That is, they are often used to provide learners with visual feedback on their performance, raising their awareness of features of their pronunciation and showing them how their pronunciation differs from that of the model. As opposed to subjective evaluations made by the teacher in old language labs or students' own comparisons between their recordings and those of a model stored, this technology allows for an objective and automatic evaluation of learners' speech (Molholt, 1988). Speech visualisation technology offers learners 'tangible' representations of speech (Anderson-Hsieh, 1992), both from their own speech as well as from that of native speakers. Visual representations of speech allow us 'to transpose a specific auditory gesture common to a closed linguistic community into a visual gesture capable of being decoded by a universal semiotic community' (Léon & Martin, 1972, p. 143). Although in this particular example Léon and Martin refer to intonation, the same applies to acoustic representations of speech in, for example, a spectrogram. In other words, it is a language-independent way of representing speech.

The tools and techniques employed for the visual enhancement of the auditory signal were originally conceived for phonetic analysis, although they were soon applied to help the hearing impaired acquire more natural speech (e.g. Abberton & Fourcin, 1975) and more recently, to train foreign language pronunciation. Visualisation software allows teachers to create their own activities, being able to design resources that cater for their students' needs depending on their individual difficulties and L1s. It is hypothesised that the additional information offered by spectrograms, pitch contours or waveforms should make learners aware of the divergences between their productions and those of the native models stored, hopefully improving the learners' attempts to match the users' visual representations to the ones by the model (see e.g. O'Brien, 2006; Pennington & Esling, 1996). The next sectoion explores the different types of display that have been used in CAPT and some of the results obtained.

Pitch contours

Displays showing pitch contours have been commonly used to teach intonation. They make suprasegmental features more 'tangible' (Anderson-Hsieh, 1992) by offering rather straightforward representations that are 'relatively iconic, with rising, falling, and level lines on the display usually corresponding to rises, levels, and falls in a speaker's voice pitch' (Levis, 2007, p. 190). Although the use of these techniques has not always brought about the expected benefits (e.g. Vardanian, 1964; Wichern & Boves, 1980), perhaps due to the level of detail provided by the display (Cranen, Weltens, de Bot, van Rossum, 1984), its use to facilitate the acquisition of prosody has been long advocated (Anderson-Hsieh, 1992, 1994; Chun, 1998; Léon & Martin, 1972; Spaai & Hermes, 1993), with a good number of studies offering empirical support of their usefulness for facilitating FL intonation training (de Bot, 1983; de Bot & Mailfert, 1982; Hardison, 2004, 2005; James, 1976, 1977; Ramírez-Verdugo, 2006; Taniguchi & Abberton, 1999).

James (1976) found that by simply showing a pitch contour of the audio model to imitate was in no way better than merely listening to a model and repeating after it, whereas offering visual feedback on the learners' intonation yielded significant improvements. Thus, the key aspect in the visualisation of intonation lies, as mentioned above, not only on the visual representation of the model's intonation, but on the possibility to offer feedback concerning the learners' productions, which may then be compared to the model sentence.

The study by de Bot (1983) offers evidence in favour of using audiovisual feedback to teach FL intonation, as a group receiving audio-visual feedback obtained much better results than a group receiving only auditory feedback. Furthermore, he found that the group receiving audio-visual feedback tended to make more attempts at improving their pronunciation than those receiving auditory feedback alone. He postulates that the former decided to repeat sentences or not depending on what they saw, whereas the latter relied on their own perceptions of their attempts. This is particularly problematic given the learners' 'phonological deafness' (Renard, 1979) rendering them unable to evaluate their own performance. Learners may assume they are pronouncing something well when they are not, as they will not be able to perceive subtle differences between the phonology of their L1 and that of the FL. As pointed out by de Bot (1983), learners who received visual feedback could have been more motivated and tried harder to correct the error, but it could also be that they simply did not know that their pronunciation was wrong (or why).[31]

In a similar line, Hardison (2004) employed visual feedback in order to train FL prosody, but instead of using the native speakers' pitch contour as a model to imitate, she used it as feedback after learners' attempts. Hardison claims that this would allow learners to gain more confidence in their ability to produce sentences in the FL and, thus, have potential for generalisation to unscripted speech. Her data reveal that the experimental group obtained significant improvements from training, including generalisation to segmental accuracy and to novel sentences.

Participants in Ramírez-Verdugo's (2006) study became more aware of the functions of intonation for communication and of the prosodic organisation of speech after receiving explicit visual feedback on their pronunciation of the FL intonation, improving, in turn, their production of the FL intonation too. She found that in the pre-test, learners

---

[31] de Bot (1983) adds a word of caution and claims that improvements in the group that received training could be due to the mere fact of devoting more time to pronunciation practice. However, the experimental and control groups in Taniguchi and Abberton's (1999) study received the same amount of explicit pronunciation training (use of visual feedback being the independent variable) and the control group was clearly outperformed by the experimental one.

frequently employed the same (limited) patterns, often due to overgeneralisation of a level falling contour and/or transfer from their L1. However, she reports that post-test productions were judged as more similar to those of native speakers in terms of form and pragmatic meaning, also showing a much wider variety of tones (see Figure 2 for an example of the visual displays employed by Ramírez-Verdugo, 2006).

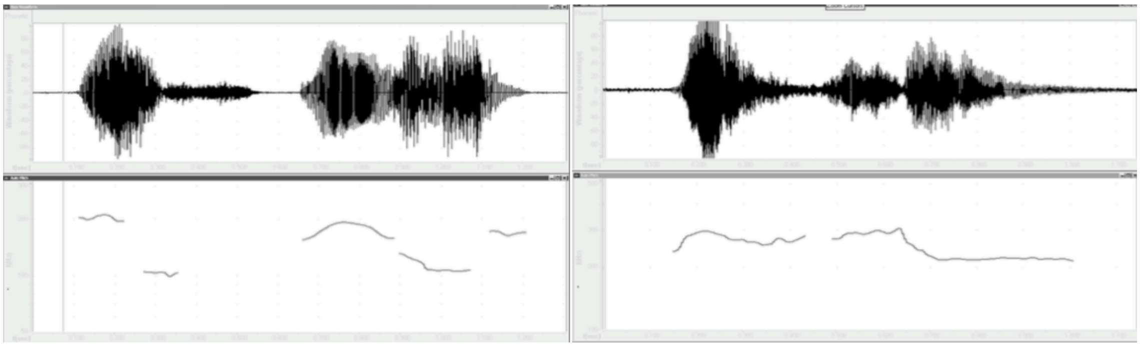


Figure 2. Example of visual feedback on a learner's intonation for the answer 'Yes, usually'. Native speaker model (on the left) and learner's production (on the left). Taken from Ramírez-Verdugo (2006, p. 152)

Despite the positive results obtained by these studies, several authors advocate an approach that deals with discourse-level intonation, rather than sentence-level practice, as the former does not fully convey the attitudinal meanings of intonation that can be grasped from discourse-level communication (Avery & Ehrlich, 1992; Chun, 1998; Levis & Pickering, 2004; Hardison, 2005). Levis and Pickering (2004) showed that sentence-level intonation interferes with discourse-level intonation by obscuring meaning can only be understood in a larger context. They claim that the traditional approach for intonation practice (i.e. sentence reading) does not encourage engagement in meaningful language use, since real life communication does not happen in isolated sentences, but in context.[32] However, analysing running speech with pitch contours is rather challenging, as sentences need to first be recorded on the computer and then processed. This is the reason why recent studies have explored new ways of illustrating intonation in live speech.

For example, in order to improve the liveliness of learners speech in oral presentations, Hincks and Edlund (2009) have recommended the use of flashing lights showing the amount of pitch variation speakers produce (Figure 3). Under the assumption

[32] As a case in point, Hardison (2005) explored whether training with sentence-level materials (a limited context) could generalise to discourse-level materials (the context where learners will ultimately need to communicate), and whether training with discourse-level materials could yield better results in natural discourse-level. Her results show that, besides a general improvement in prosody towards native-like proficiency by using visualisation technology, discourse-level practice resulted in a better transfer of this improvement to novel natural discourse.

that pitch curves may be difficult to interpret and its imitation problematic or undesirable for some learners, they propose a new way of showing learners when they are being 'monotone' in oral presentations. This technique allows analysing pitch variation in running speech, therefore offering the possibility to move beyond practice with sentences in isolation. In their study, participants increased the amount of pitch variation used in their oral presentations and this improvement was generalised to their pronunciation in new, unrehearsed oral presentations.

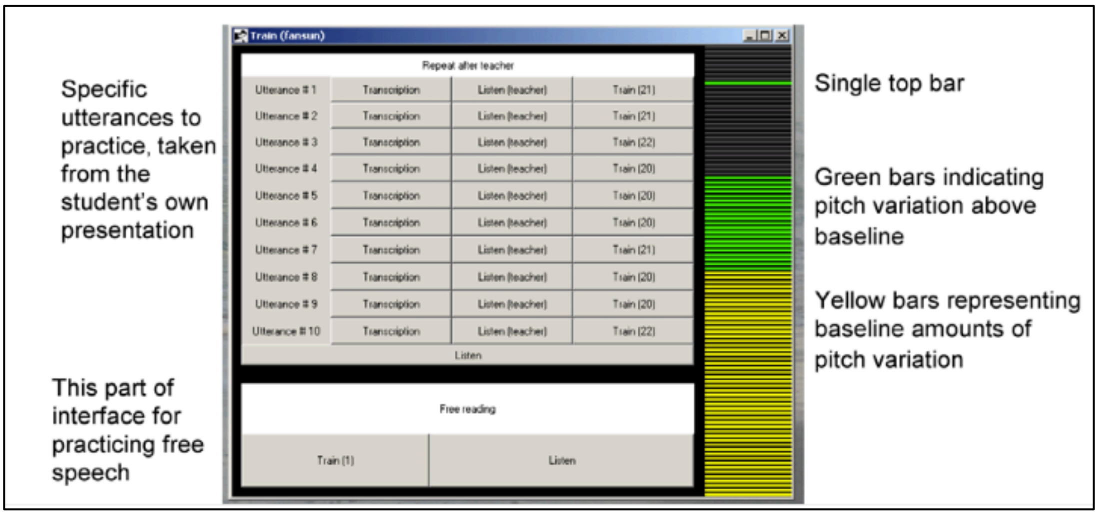


Figure 3. Example of the training interface used by Hincks and Edlund (2009, p. 38)

Waveforms

The use of waveforms has been commonly criticised when implemented in courseware for autonomous practice because of the little informativeness of the display for the average language learner (see Llisterri, 2001, 2007). Nevertheless, several researchers have evaluated the usefulness of these displays in order to illustrate certain features of pronunciation. Motohashi-Saigo and Hardison (2009) used waveforms to display the duration of problematic segments in order to help English learners improve their perception and production of Japanese geminates. They found that learners' perception improved after training and this improvement was transferred to production and generalised to novel stimuli. They recommend this type of tool as a supplement to classroom instruction, as it produced positive outcomes and it was well received amongst participants. They state that, with minimal instruction, learners were able to focus their attention on the problematic aspects and improve both their perception and production.

In a similar vein, Hew and Ohki (2004) employed waveforms and pitch contour displays in order to offer visual feedback on the pronunciation of a group of Malaysian learners of Japanese (see Figure 4). Although the main goal of the study was to investigate the effects of animated graphic annotations on the acquisition of a range of problematic aspects of Japanese pronunciation, they found that learners receiving feedback with waveforms and pitch displays also improved their pronunciation considerably. Both treatment groups, the one receiving immediate visual feedback with waveforms and pitch contour displays and the one trained with animated annotations, performed substantially better than a group receiving auditory feedback alone. Their findings offer further suppor to de Bot's (1983) observation that students in the visual feedback group repeated their words more often and made further efforts to improve their pronunciation, possibly because they were offered external objective information about their pronunciation.

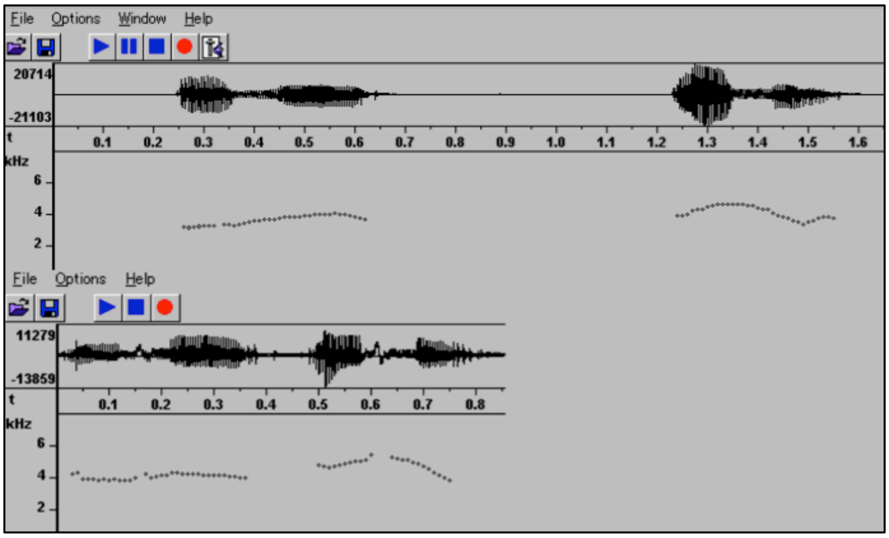


Figure 4. Example of the waveforms employed by Hew and Ohki (2004, pp. 401-402). Model native speaker (on top) and student's production (below)

Spectrograms

Numerous authors have also advocated the use of spectrographic displays as a tool to help learners detect their pronunciation mistakes (Akahane-Yamada et al. 1997; Lambacher, 1999; Molholt, 1988, 1990; Molholt & Hwu, 2008; Olson, 2014). As Molholt (1988) notes, this type of display offers objective measures of learners' pronunciation showing the exact features that need to be changed (e.g. formant frequencies, aspiration, duration, or word and sentence stress). In his words, 'because the communication is in the visual mode, students do not need to learn the linguistic vocabulary associated with phonological analysis' (Molholt, 1988, p. 92). He claims that this type of display will show learners the exact features that have to be modified, by pinpointing the exact location of mistakes as well as the progress made in subsequent attempts to correct them. The underlying assumption, in line with Léon and Martin's (1972) conception of visual feedback as some kind of universal language mentioned above, is that, with the necessary practice, learners will be able to associate the representations shown on the screen with the movement of their articulators (see Molholt, 1988, 1990; Lambacher, 1999).

Akahane-Yamada et al. (1997) investigated the effectiveness of spectrographic feedback on the pronunciation of English /r/ and /l/ by a Japanese speaker. Despite having only one subject, they report substantial benefits in the production of these two problematic sounds, with the subject's intelligibility scores improving from 44.1% in the pretest to 97.1% in the post-test after only five hours of training – results that the authors consider much more positive than those of other studies that have fostered perceptual training alone. This findings were replicated in a subsequent study with 10 subjects, where intelligibility scores improved from 62.7% to 85.1% and goodness rating scores from 3.36 to 4.18, showing also a slight improvement in perception accuracy from pre- to post test (Akahane-Yamada et al., 1998). In a similar line, Ruellot (2011) used spectrographic feedback to illustrate the pronunciation of French /u/ and /y/ to English native speakers. She reports that subjects improved their pronunciation of the two phonemes after training, but she did not find any additional improvements from working with audio-visual feedback as compared to receiving only audio feedback. In a recent study, Olson (2014) tested the effectiveness of spectrographic representations coupled with waveform displays in PRAAT in order to teach English learners of Spanish not to produce intervocalic /b/, /d/, and /g/ as plosives, but as their approximant allophones [β], [ð], and [ɣ]. He found that learners improved considerably their pronunciation of these sounds after training, that these gains were retained over time, and that benefits obtained were generalised from words in isolation to words in connected speech.

Notwithstanding the positive results reported above, it should be noted that the convenience of using of spectrograms can vary considerably depending on the aspect to be taught. Molholt (1988, pp. 97-98) argues that 'it may be very difficult for students to hear the difference at first, but they definitely can feel it and see it on the screen. […] Gradually they learn to associate the feelings with the sounds, so that they make the sounds more naturally in continuous speech after 4-6 weeks of class, 1 hour per week'. However, it could be argued that 'feeling' how modifications in pronunciation affect the resulting representation on the spectrogram will not always be easy for students. In Olson's (2014) study, for example, both spectrograms and waveforms do show very clearly the differences between stops and approximants (Figure 5); but the same cannot be said about some of the aspects taught by, for example Lambacher (1999) or Ruellot (2011) – see Figure 6 for Lambacher and Figure 7 for Ruellot.

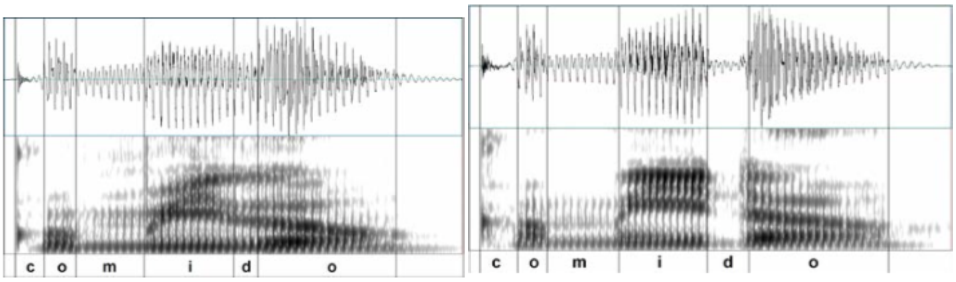


Figure 5. Spectrographic and waveform representation of the word *comido* as pronounced by a Spanish (on the left) and an English (on the right) speaker. Taken from Olson (2014, pp. 179-180)

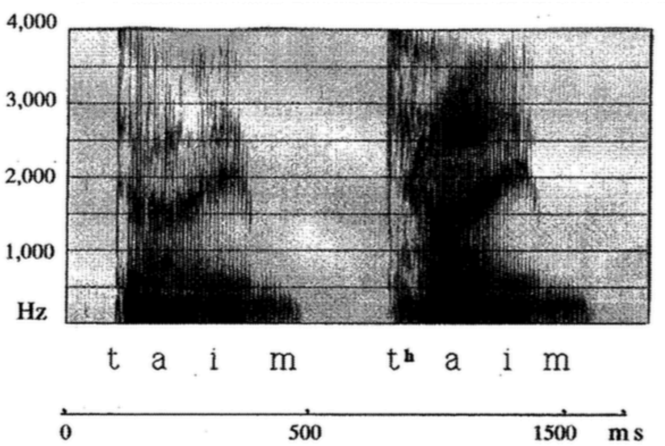


Figure 6. Spectrograms of the word *time* showing the differences in aspiration between [t] and [t^h]. Taken from Lambacher (1999, p. 144)

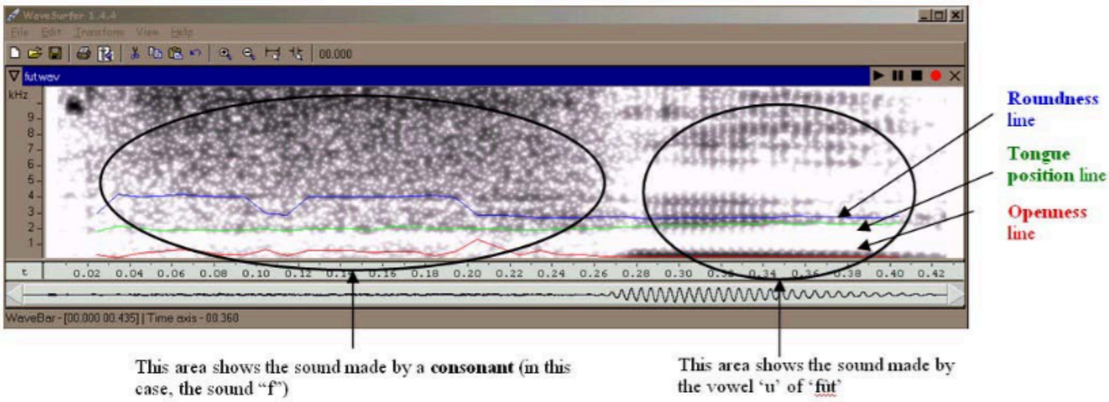


Figure 7. Spectrographic representation used to teach English learners to pronounce /u/ and /y/ in French. Taken from Ruellot (2011, p. 201)

Some of the recommendations given by Lambacher (1999) will undoubtedly require some kind of explanation. Figure 6 illustrates one way of teaching the difference between an aspirated and an unaspirated /t/. Looking at Figure 6 through the eyes of a general language learner, it seems reasonable to expect that in order to understand that a burst of air represented by a vertical line followed by absence of energy shows aspiration in a spectrogram and explain how to change it, aspiration should be defined first. Thus, learners would need, on the one hand, certain metalinguistic vocabulary to understand certain concepts, and on the other hand, linguistic notions and skills in order to interpret the spectrographic feedback.

The approach recommended by Lambacher contemplates spectrograms as an additional activity to be used with the guidance of a teacher, not as a stand-alone training method for autonomous practice. Students would first interpret the sound patterns on the screens, then practise the target sounds in sentences and dialogues, and finally try to imitate a target model with the guidance of a teacher. Although they may come in handy to illustrate very problematic fossilised pronunciations that learners cannot perceive in any other way, using this kind of representation does not seem very suitable or practical for the average language learner.

Wilson (2008) suggests using this kind of technology in combination with Moodle in order to make a better use of classtime. The approach he recommends consists in learners paying attention to certain values of their speech, as offered by a spectrographic representation, and then selecting among a range of values set by the teacher. In this way, the teacher can monitor learners' performance and offer them individualised feedback. However, the approach seems rather laborious for the teacher and not very easy for students either. Therefore, despite the potential of this type of display for drawing the learners' attention to problematic sounds, its implementation does not seem very feasible as, in the end, a general FL teacher would be forced to teach phonetics in a FL course where it was already difficult to find time for pronunciation. This would impose extra demands on the teacher's knowledge of pronunciation, in this case, acoustic phonetics, something which seems especially problematic given that one of the reasons why pronunciation teaching is neglected is the fact that teachers do not feel prepared to teach it (see section 1.1.3).

Other visual information

Because of the lack of transparency of certain types of visual representations of speech (e.g. spectrograms), some authors have proposed innovative alternatives with different types of visual information. For example, Gómez et al. (2008) recommend using videogame interfaces that offer simplified versions of spectrographic feedback. They recommend game-like animations showing the degree of correctness in the learners' pronunciation after performing a 'behind the scenes' evaluation obtained by matching the students' pronunciation to a model stored in the computer. Animations may take different forms, like darts that get closer to the center of a dartboard as learners' attempts get closer to the model stored or cars that follow an adequate trajectory when students pronounce correctly but which deviate as learners' attempts drift away from the model, for example (Figure 8).

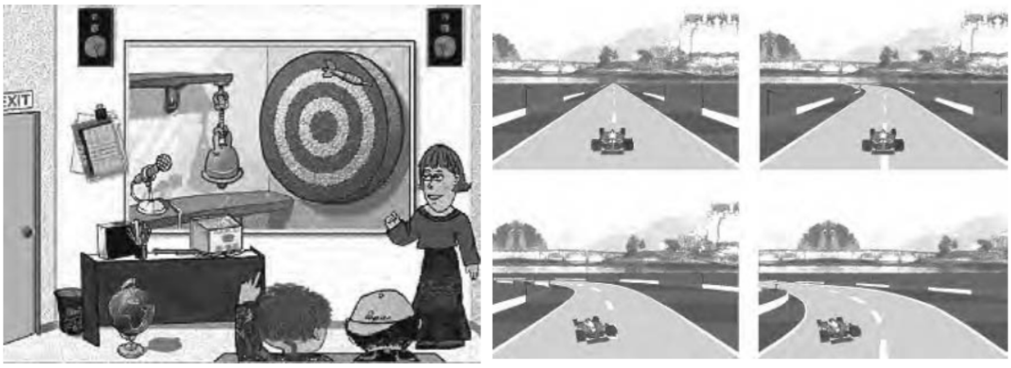


Figure 8. Examples of game-like feedback recommended by Gómez et al. (2008, pp. 61-62)

As commented above, Hew and Ohki (2001, 2004) explored the effects of animated graphic annotations [33] in the acquisition of several problematic segmental and suprasegmental aspects of Japanese pronunciation. In their 2001 study, the researchers found that learners improved in terms of listening ability but not their pronunciation, whereas in their 2004 study, participants in the experimental group improved their pronunciation significantly. In this later study, they added an additional type of training with visual feedback (using waveforms and pitch contour displays) and found that students in the animated graphic annotations group benefitted significantly from instruction as compared to the control group (using only text and audio). However, there were no significant differences between the group receiving instruction with graphic annotations and the group using waveforms and pitch contours.

Dowd, Smith, and Wolfe (1998) investigated the effectiveness of visual feedback that showing spectra representations of resonance frequencies of the vocal tract. Feedback was given in the form of two vertical lines showing the resonance frequencies that had to be matched to those of a model stored. The lines illustrated the frequency measures of the first two resonances of the vocal tract, which were used to measure accuracy in the pronunciation of French vowels. The group that followed instruction with visual feedback obtained much better results than a control group following traditional training.

[33] Symbols designed to help learners of Japanese to improve their listening and pronunciation skills by visually illustrating differences in features such as pitch, voice or length.

Other studies have offered visual feedback based on the learners' tongue position during articulation. For example, Flege (1989) used a glossometer that provided visual feedback on tongue position when articulating vowels. He found that, with only three days of training, the subject improved her pronunciation between /i/ and /ɪ/ by receving feedback on how close her pronunciations were to a target articulation. Although the study involved only one subject and the author acknowledges a number of limitations, Flege's data indicate that the subject was able to interpret visual information about tongue position and to implement it into her speech.

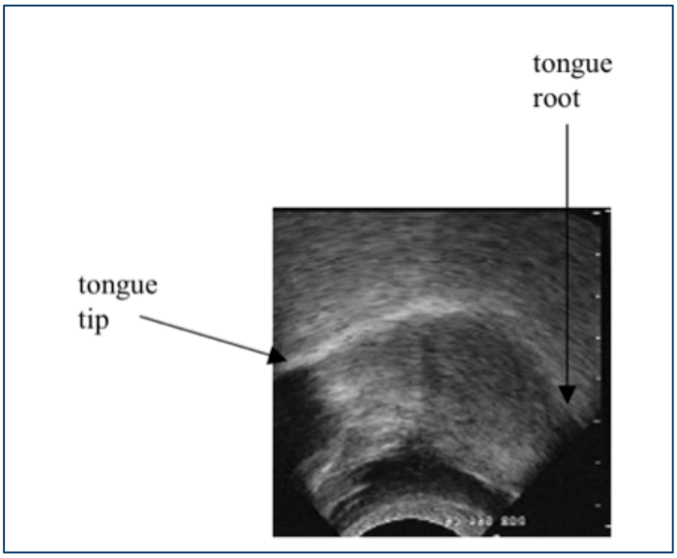


Figure 9. Example of ultrasound image to illustrate tongue articulation used by Gick et al. (2008, p. 310)

Gick et al. (2008) discuss the possibilities of ultrasound imaging showing two-dimensional images of tongue position during articulation (Figure 9). They claim that this technique circumvents some of the obstacles posed by similar technologies, given that it is affordable, non-invasive, portable and quick. They comment on positive findings of similar studies with native speakers with delayed mastery of /r/. Besides, they report beneficial effects of ultrasound imaging in FL learning after only 30 minutes of training. In their pilot study, participants were able to improve their pronunciation of problematic sounds in contexts that posed difficulty and to generalise gains to words that were not included in the training.

After a revision of available software offering visual feedback for vowel training, Carey (2004) used Sona-Match for an empirical test with Korean learners of Australian English. Learners were presented with visual displays of their vowel space and video presentations of mouth movements for the target sound /æ/ as well as for other L1 sounds that are commonly transferred as a result of interference (Figure 10). He found that after only 5 hours of instruction, participants in the experimental group improved their production of the target vowel. However, significance was only reached for vowels produced in citation form, not for continuous forms.

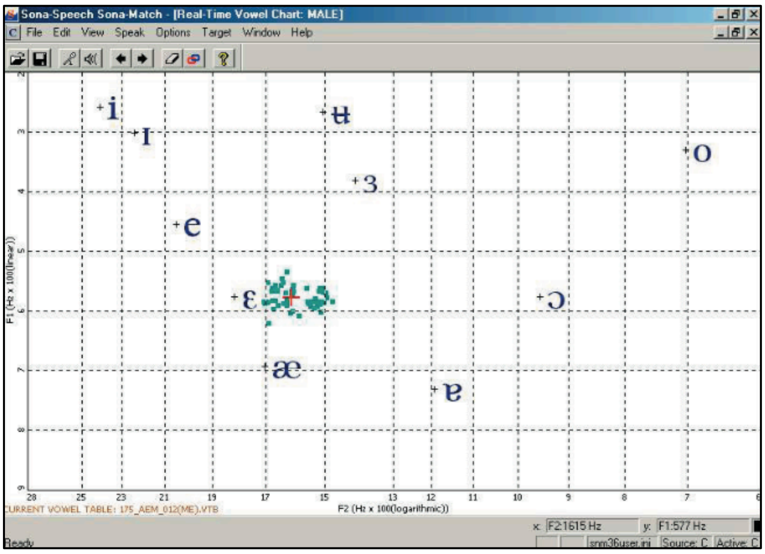


Figure 10. Visual display used by Carey (2004, p. 10) to offer feedback on vowel productions

Finally, as an alternative to commercial software that measures acoustic properties of speech, Brett (2004) advocates the use of Praat (Boersma & Weenik, 2015) to help learners improve the pronunciation of foreign language vowels. He suggests offering feedback by measuring formant data from learners' output and plotting it on a graph that exemplifies articulatory information in a manner similar to the one provided by the vowel quadrilateral. Although the technique seems potentially advantageous for offering automatic feedback on such a problematic aspect of FL phonology, he mentions a number of issues that may limit the usefulness of the approach, such as background noise, phonetic contexts that may alter the acoustic properties of the vowel, the slightly delayed feedback, or the fact that formant values used as targets are relative, not absolute values. Even though he highlights that Praat can be adapted to different needs given its open source nature or the fact that it is free, he recommends elaborating a more pedagogical display, as Praat was initially designed for research purposes and may not be very easy to understand for the average learner.

Some concerns & directions

The aforementioned methods are potentially promising for CAPT in that they allow learners to receive individualised, private, limitless practice on very specific aspects of their pronunciation. They offer objective feedback on learners' pronunciation and they do so through the visual mode, different from the often-biased perception learners experience through the auditory channel. As suggested above, learners could work on problematic aspects of their pronunciation outside the classroom in order to automatise adequate productions of the FL phonology and then, be able to put those aspects into communicative practice in the classroom. However, these methods are not free from limitations.

One of the most cited problems with this type of feedback is the difficult interpretability for learners (Carey, 2004; Hincks, 2015; Neri et al., 2002; Levis, 2007; O'Brien, 2006, Pi-Hua, 2006). Even though some visual displays in current software have been adapted for language teaching and offer representations that could be considered to be much more pedagogical in nature,[34] the fact that these techniques were initially devised for phonetic research purposes makes them, in some cases, extremely difficult to interpret. Modifying articulatory gestures from acoustic representations is not an easy matter, as 'there is often no simple correspondence between gesture and acoustic structure' (Akahane-Yamada et al., 1997, p. 341).

Pitch contour representations of intonation could be considered to be among the easiest visuals to interpret. As Levis (2007) points out, they are 'relatively iconic, with rising, falling, and level lines on the display usually corresponding to rises, levels, and falls in a speaker's voice pitch' (pp. 190-191). Nonetheless, they impose limitations on the type of materials that can be used. For instance, the presence of unvoiced segments when illustrating intonation will not yield full pitch contours and may require additional explanations, therefore not making information presented to learners easily interpretable and possibly requiring additional explicit explanations (Anderson-Hsieh, 1994).

Most of the approaches mentioned require expert knowledge and may pose difficulties even for teachers who are not familiar with phonetics. In fact, this was one of the claims by the students evaluating the MyET software in the study by Pi-Hua (2006) with regard to spectrograms and waveforms. They stated that it was impossible for them to understand what was wrong about their pronunciation by simply looking at the information offered by these displays, without receiving explicit explanations on how they worked. Since one of the main reasons why pronunciation teaching was neglected was the teachers' lack of confidence and training to teach pronunciation (see section 1.1.3), unless simplified, user-friendly versions of the above-mentioned technologies are used,

[34] See Llisterri (2001, 2007) for a review.

implementing them on a regular language classroom or expecting learners to use them autonomously at home seems rather unfeasible.

Despite the fact that some consider these kinds of representations intuitive and argue that students could 'feel' how modifications in their pronunciation will modify the represented pattern (Molholt, 1988, 1990), the visual feedback offered by some of these methods does not explicitly inform learners about the nature of their mistakes or how to correct them, at least not in an obvious, transparent manner (Neri et al., 2002). Thus, as Neri et al. (2002) observe, learners' random repeated attempts at improving 'the presumed errors' may in some cases lead to fossilisation, rather than improvements.

Additionally, as mentioned above, merely trying to match our production to a model's pronunciation will not be a desirable goal in some cases. As Neri et al. (2002) write, 'the fact that the system shows two comparable displays […] wrongly suggests that the ultimate aim of pronunciation training is to produce an utterance whose spectrogram or waveform closely corresponds to that of the model utterance' (p. 453), which somehow also implies that the ultimate goal of pronunciation is to sound like a native speaker. Furthermore, as exemplified in Llisterri (2001, 2007),[35] different representations of the same word may be perfectly correct and intelligible, in some cases, being even impossible for the same speaker to obtain the same representation of an utterance twice (Carey, 2004). This could lead to frustrating and counterproductive experiences if learners waste time trying to match a model when their pronunciation is already correct.

Methodologically speaking, it would be interesting to explore alternative methods to help learners attain new targets of the FL phonology. As Carey (2004) recommends, when exemplifying sound contrasts, training should also address contrasts between the learners' L1 and the FL, as opposed to traditional segmental pronunciation training focusing exclusively on contrasting minimal pairs in the foreign language. It is true that learners may find it difficult to discriminate and produce similar sounds in the FL if they only differ subtly from one another given that learners are often unable to perceive differences that are not distinctive in their L1. However, using the learners' L1 as a reference point may be extremely useful, since ultimately, learners will need to know how to modify their pronunciation from that reference point to a target model, rather than distinguishing

[35] Llisterri (2001, 2007) illustrates this with two waveforms showing two different representations of the word 'hola', as pronounced by two speakers. He claims that despite clear differences in duration and amplitude in the waveform representations of the two realisations, this does not imply that any of the two should pose comprehension problems for a native speaker (see Llisterri, 2001, pp. 11-12).

between two FL sounds that may be completely new to them. Departing from the articulation of sounds students are already familiar with may be a facilitative technique (Molholt & Hwu, 2008), as in some cases learners will only need to modify their articulatory movements slightly from what they commonly articulate for their L1 sounds (see e.g. Flege, 1989).

For these methods to be beneficial to the average language learner, the relationship between technology and pedagogy must be conscientiously addressed (Neri et al., 2002; Pennington, 1999; Pi-Hua, 2006). So far, most of the studies evaluating the effectiveness of certain software have been conducted in extremely controlled, laboratory-like environments and for very short periods of time. For the real advantages of CAPT to be fully exploited, learners should be able to use the technology at home, where they feel relaxed and comfortable, without the inhibiting presence of others or the time constraints of the classroom. One of the advantages of CAPT mentioned above is that it could offer the extra time students need, something often difficult in classrooms with many students.

When applied to the FL classroom, this is true to the extent that learners can have more individualised instruction as compared to having only one teacher for the whole class. Each student may have a 'teacher' that works exclusively with him/her (the computer). Nevertheless, the class will get to an end, and students may have not had the necessary practice. Limiting pronunciation practice to a language lab/computer lab is only slightly better than an old language laboratory. Current technology is much more advanced and offers numerous possibilities, but when the use of technology has to be limited to the classroom, the time learners can devote to pronunciation practice is still up to the teacher, not to them. Besides, if learners need to record themselves in the language lab, the environment will certainly be less imposing than the class, where everyone is listening to the student the teacher is monitoring (only one teacher, only one student talking). Nonetheless, having a classmate half a meter away when you need to record yourself is not precisely a 'stress-free environment' either, as students are not really on their own. Learners should be able to practise at home, whenever they want and whenever they need.

A possible solution to this may come from whole-tutoring systems. These are often commercial software incorporating several of the above-discussed technologies for pronunciation teaching. For example, learners are given sentences to practice and are offered feedback on their performance by means of ASR feedback. Feedback may range from a simple score indicating how well the user did, to visual representations of the students' speech pinpointing the exact location of the learners' mistakes (by means of waveforms, pitch contours, etc.). As noted at the beginning of this section, since the purpose of this chapter is to revise the ways in which technology can enhance FL pronunciation training, specific software will not be dealt with here, as the underlying technology (with its advantages and limitations) is the same technology that has already been explored.[36]

Visual enhancement of sound articulation: Talking heads

A very original approach in computer-assisted instruction is the implementation of animated agents that simulate the presence of a human teacher in the computer. These animated characters offer students the virtual presence of a tutor that guides them through their learning process and provides immediate feedback on their performance. Like other CALL tools, automatic tutors offer limitless patience. The tutor does not become angry, tired or bored, and it is available 24 hours a day, seven days a week. This, as Massaro (2006) notes, is especially advantageous for children with special needs, such as autistic children with irregular sleep patterns. Since they create the illusion of life, talking heads are well received by students, something that could increase motivation and make students more engaged with educational software (Lester et al., 1997).

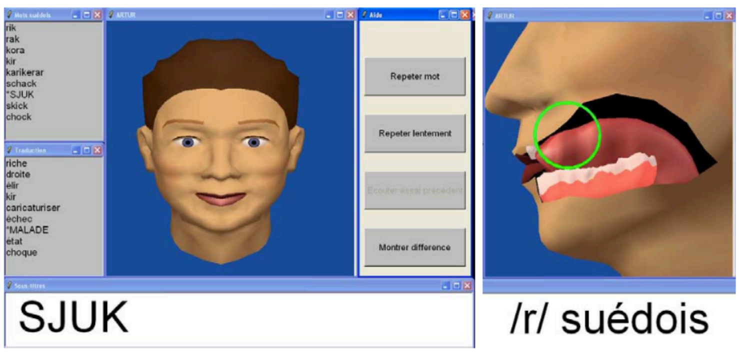
The role of multimedia learning has been commonly emphasised given that it facilitates the assimilation of information and caters for different learning styles (Mayer, 1997; Mayer & Moreno, 2002). However, talking heads do not only offer a more motivating, multi-media environment, they also transmit information not easily obtained from the acoustic signal alone. Even though visual articulatory information is not only informative for FL learners, as we will see below, native speakers do not need to see where articulators are placed in order to produce sounds adequately because they have acquired the phonological system of their L1 naturally. Nevertheless, for the adult FL learner who suffers the effects of 'equivalence classification' (Flege, 1987), improving the pronunciation of a sound that is very similar to a native sound can be extremely difficult without explicit information on articulation.[37]

[36] For a review of some of these programmes, see Llisterri (2001, 2007).

[37] Talking heads can help learners realise subtle nuances in sounds learners assume to be similar or the same as they have in their mother tongue, when they are not. For example, the differences in articulation between the Spanish and English /t/ and /d/. The Spanish sounds are dental and their English counterparts alveolar. Many learners who have been speaking the language for years know that they 'sound different' but do not know why. Talking heads are a great help in this regard. Subtleties that may go unnoticed for students for years can be easily pointed out by illustrating articulation through animations.

Animated agents employed for pronunciation training usually take the form of talking heads (Figure 11).[38] These animated faces are commonly used to illustrate the articulation of sounds by showing how the different articulators move (the lips, tongue, teeth, or even vocal fold vibration) in order to produce those sounds, an often-recommended technique for pronunciation training (Baker, 2006; Catford & Pisoni, 1970; Hancock, 2003). These faces are undeniably appealing for pronunciation instruction, especially because they can illustrate aspects of the FL phonology without requiring explicit knowledge of phonetics.

Both synthetic speech and natural speech can be used with these faces, although neither of them is free from limitations. As noted above, using synthetic speech is subject to empirical validation as regards its appropriateness for the acquisition of FL phonology, whereas using natural speech would limit the simplicity of creating lessons, as content should be recorded in advance (Massaro et al., 2012), and non-native teachers would need to resort to native speaker models. Additionally, some of these agents also incorporate a speech recognition module that detects learners' mispronunciations and offers feedback, although as will be discussed below, current technology does not meet the standards required for pronunciation instruction.



(a) User interface (b) Articulatory illustration

Figure 11. Example of Engwall and colleagues' virtual teacher (ARTUR) showing a front view of the face (a) and an augmented reality display showing internal articulatory movements (b). Taken from Engwall (2008, p. 2)

Talking faces hold great promise for pronunciation training, as visual information is commonly employed in speech in order to complement the auditory modality. As Hardison (1999) writes, 'in face-to-face communication, input is matched to the observers'

[38] They are also referred to as 'talking faces'; both terms will be used indistinctively here. They have been used in a wide variety of programmes for pronunciation instruction but also for general language learning, offering conversational practice through simulated dialogues with the computer (e.g. Morton & Jack, 2005 in their self-access learning package Spoken Electronic Language Learning - SPELL).

perceptual categories (phonetic and visual) for identification' (p. 218). It is well known that people with hearing impairments make common recourse to lip-reading in order to complement the acoustic dimension of speech. However, speech is also perceived bimodally by subjects with normal hearing, as evidenced by the McGurk Effect (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976).[39] Research has shown that people often make recourse to visual information whenever they require extra information or support. A 'joined effort' of the auditory and visual channels in subjects with normal hearing has been found under difficult listening conditions (Binnie, Montgomery, & Jackson, 1974),[40] as visual cues make speech more robust to environmental degradation (Hazan, Sennema, Iba, & Faulkner, 2005). This reliance on the visual channel has also been found in FL learners, who pay more attention to visual information when presented with foreign language stimuli (Hazan & Li, 2008; Sekiyama & Tohkura, 1993).

A number of studies have explored the role of talking heads in the facilitation of speech perception. Visual cues offered by animated faces have proven to increase intelligibility of the speech signal (Siciliano, Faulkner, & Williams, 2003), improve identification of problematic sound contrasts for people with hearing impairments (Massaro & Light, 2004), and help learners distinguish difficult L2 contrasts (e.g. Hardison, 2003; Hazan et al., 2005; Ortega-Llebaria, Faulkner, & Hazan, 2001). Nevertheless, if only frontal views of the speaker's face are used, the degree of informativeness of visual cues will depend on the number of visemes[41] in the language (Hazan et al., 2005). As a case in point, Hazan et al.'s (2005) results on a lip reading test with a group of native speakers showed that identification was much higher for the contrast between the labial /b/ and /p/ and the labio-dental /v/ (92%), than the contrast between /l/ and /r/ (72.1%), given that the former contrast was visually much more salient.

Research suggests that L2 learners do not make use of visual cues that are not contrastive in their L1, despite clear visible differences in articulation (Ortega-Llebaria et

[39] The McGurk effect is a perceptual phenomenon that exemplifies the interaction between two different modes of speech perception, namely hearing and vision. McGurk and MacDonald (1976) investigated the impact visual information had in the perception of the English syllables /pa/, /ba/, /ka/, /ga/. This effect is commonly explained by the perceptual illusion that occurs when a combination of an audio stimulus of, for example, a bilabial articulation (e.g. /ba/) and a video stimulus of a velar articulation (e.g. /ga/) is perceived as an intermediate position, that is, as alveolar (e.g. /da/). Participants in their study identified the different syllable conditions accurately in 99% of the cases when they were only exposed to the audio stimuli. However, when the audio /pa/ was contradicted with a visual /ka/, 81% of participants perceived it as /ta/.

[40] The reader is referred to Hardison (1999, 2007) for an extensive review of the literature on the impact of vision on the perception of speech in native and non-native speakers.

[41] A term analogous to the concept of 'phoneme', it refers to sound categories that can be discriminated visually (Massaro, 1987, p. 36).

al., 2001). Nonetheless, it can be hypothesised that provided learners were trained to pay attention to visual cues, interior views of mouth movements might be extremely useful in order to reveal subtleties in articulation. For example, most EFL Spanish learners without explicit pronunciation instruction will probably pronounce English /ʃ/ as the /s/ as a result of their L1 interference. Although an eventual correction of this pronunciation by mere exposure cannot be discarded, it can be argued that most adult students will notice something different in their pronunciation but will not know exactly what or why. Hence, providing learners with visual articulatory information should help them attain a more accurate production of the target sound.

Some have argued that methods that encourage the emulation of tongue positions by simply looking at internal movements of articulations are flawed, given that learners cannot 'see inside their heads to make comparisons' (Carey, 2004, p. 5). However, while this may be more challenging for vowels given the lack of concrete places of articulation, learners could benefit from articulatory information for the pronunciation of consonants. Despite the fact that representations of articulatory movements usually offer canonical representations of sounds articulated in isolation (Llisterri, 2007), it can be hypothesised that by becoming aware of where a sound should be pronounced (by visually imagining it and perhaps silently articulating it) learners should be better equipped to modify incorrect articulations of the sound in question and eventually, even facilitate the establishment of a new phonetic category.

Talking heads have long been used as an aid for the articulation and perception of foreign language sounds (Engwall, 2008; Engwall & Bälter, 2007; Hardison, 2003; Hazan et al., 2005; Liu, Massaro, Chen, Chan & Perfetti, 2007; Massaro & Light, 2003; Massaro, Bigler, Chen, Perlman, & Ouni 2008; Wik & Hjalmarsson, 2009), for children with hearing (Massaro & Light, 2004) or articulatory (Fagel & Madany, 2008) difficulties, or even to enhance telephone conversations to help the hearing-impaired (Siciliano et al., 2003). One of the most well-known talking faces is Baldi, created by Massaro and his colleagues and which now has different versions for Italian, Arabic, Chinese, Spanish (see Massaro, 1998). They have created a canonical face and then reshaped it to match different facial models so that learners feel identified with the animated tutor, by having a tutor that is more similar to them (Massaro, Liu, Chen, & Perfetti, 2006).

Several studies have found positive effects of training learners with a front view of a talking face. For example, Hardison (2003) found that students from two different language backgrounds (Korean and Japanese) receiving audiovisual training on the perception of problematic sound contrasts obtained significantly greater improvements in perceptual accuracy as compared to groups receiving training with audio alone. Moreover, the gains obtained thanks to perceptual training were also transferred to a better production of the sounds without explicit pronunciation instruction. Similar results were obtained by Hazan et al. (2005), who found that training in perception also resulted in better production, even for the less visually distinctive contrast mentioned above (/r/-/l/).

Nevertheless, one of the most promising aspects of talking heads lies in the possibility to offer visual instructions on how to place articulators in order to produce FL segments, especially on those sounds that are not visually perceivable from an ordinary view of a person's face. Massaro and Light (2003) investigated the use of their talking head to improve Japanese speakers' perception and production of the /r/-/l/ contrast by showing an internal view of the articulators. Even though learners improved in both identification and production of the contrast, despite having found effects of generalisation to novel stimuli, and even though learners preferred instruction with the talking head, the differences between using a normal (frontal) view of the face and a view showing the articulators were not significant. In other words, seeing internal articulatory information did not offer additional benefits.[42]

Notwithstanding the above, the results obtained by Badin, Tarabalka, Elisei, and Bailly (2010) suggest that learners can indeed interpret tongue movements and that perception is improved when learners can rely on this information (see Figure 12 for an example of the articulatory information offered by Badin et al., 2010). However, they found that participants paid more attention to the lips than to tongue movements, and that a normal view of the face was perceived better than a cutaway view of the head, unless auditory information was absent. Liu et al. (2007) offered further evidence supporting the use of articulatory information for pronunciation instruction. They measured the improvements in production obtained after using three different methods: audio only, a human face, and Bao (Baldi's Chinese version). Although no significant differences were found between the virtual and the human tutor, the synthetic speech and visual training showing internal articulators offered by Bao was significantly better than the audio only version. In a similar line, Engwall (2008) found beneficial effects of articulatory instructions given by the talking head to a group of French learners of Swedish. Even though the

[42] It should be noted though, that the study was not free from limitations. Several training stimuli presented a ceiling effect, there were only 11 participants in the study and there was an identification task that could be a further opportunity for learning (see Massaro & Light, 2003).

amount of improvement differed depending on the target phoneme, learners were able to improve their pronunciation of the two sounds studied (/r/ and /ɧ/) and they modified their articulation according to the instructions given by the talking head.

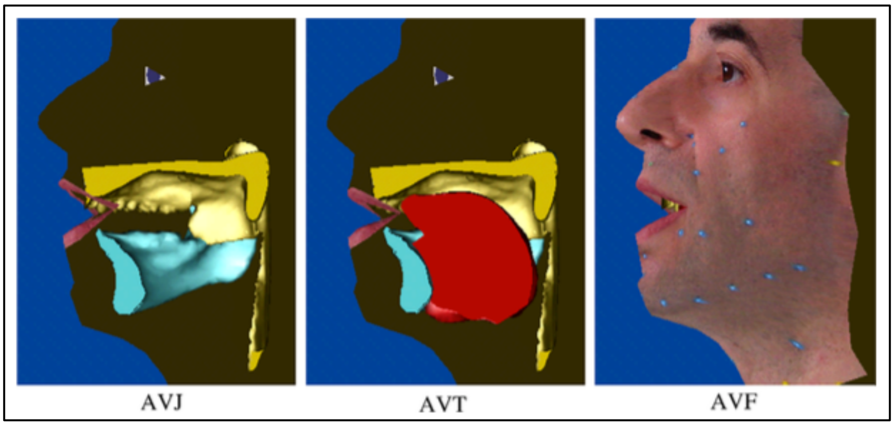


Figure 12. Example of different presentation modes in Badin et al. (2010), including cutaway view of the head without showing the tongue (AVJ), cutaway view with tongue (AVT) and synthetic face showing a complete face (AVF). Taken from Badin et al. (2010, p. 497)

Engwall (2008) envisages these animated agents as a possible solution to the problem of inadequate feedback in CAPT applications. He notes that, at best, most CAPT applications simply pinpoint pronunciation errors, without giving information on the type of error or how to correct it. Hence, talking faces offer a promising solution to this problem by illustrating how a sound must be articulated or how to modify an incorrect articulation when the system detects mispronunciations. This would be as having private pronunciation tuition with a virtual tutor that pinpoints your mistakes in a comfortable and stress-free environment. In fact, feedback in Engwall and colleagues' talking head is modelled after the kind of feedback human teachers would give and the way they would give it (see Engwall, Bälter, Öster, & Kjellström, 2006). This should be more satisfactory for learners because, as the authors claim, users tend to expect human-like behaviour from animated characters that have human appearance (Engwall & Bälter, 2007). As Lester and colleagues (1997) note, the 'persona effect' can have a strong positive impact on the learners' perception of their learning experience. These researchers found that students perceived these agents as credible, useful and entertaining, even when interacting with characters that were not very talkative and whose advisory behaviours were non-existant.

Engwall (2012) emphasises the role of talking heads for offering articulatory instructions based on the learners' mispronunciations rather than for offering general articulatory information, as he claims that self-correction may be easier than imitation. A similar approach is adopted by Wik and Hjalmarsson (2009); their system incorporates two tutors that perform different functions: *Ville* and *DEAL*. The former acts as a virtual instructor that gives feedback on students' pronunciation and language use, the latter as a role-play dialogue system for conversation training. When users make pronunciation mistakes while engaged in communication with the latter, whose main goal is to maintain an engaging conversation, they may be directed to the former for corrective feedback.

As can be seen, the advantages offered by this technology are numerous. Illustrations of mouth movements have long been used in pronunciation teaching. The Learning English section offered by the BBC World Service,[43] for example, includes videos with side and front views of a speaker who illustrates how to pronounce the different English phonemes, something that can also be found in some EFL course books (e.g. Innovations Elementary – Dellar & Walkley, 2005) and in numerous pronunciation manuals (e.g. Celce-Murcia et al., 2010; Kelly, 2000; Baker, 2006; Hancock, 2003). The fact that we can now see through skin and muscle with talking faces enhances our perception of the articulatory movements in ways that would be unthinkable (or harmful)[44] before, something potentially very promising for pronunciation instruction. Watching someone's lips is already informative enough for visemes (Hazan et al. 2005), but talking heads offer even clearer illustrations of less visible pronunciations (as the English dental articulation of /t/-/d/ mentioned above, or the post-alveolar articulation of /ʃ/ which is commonly perceived as an alveolar /s/ by Spanish speakers, for instance).

Although talking heads have proved useful in different aspects of pronunciation training (namely, in the perception and production of FL sounds), more research is needed in order to shed light on many of the aspects pertaining to the effectiveness of talking heads, as well as their optimal characteristics. For example, some studies suggest that the use of a synthetic face has not yet yielded as positive results as when using a natural face (Hazan et al., 2005; Ouni, Massaro, Cohen, Young, & Jesse, 2003; Siciliano et al., 2003), whereas in others the animated face was superior in some word positions (Liu et al., 2007), although in general, the use of a synthetic face has been shown to increase intelligibility as compared to using only audio.

Another aspect that is worth investigating is whether to use 2D or 3D animations. In Kröger, Graf-Borttscheller, and Lowit's (2008) study, for example, 2D animations were more effective in visual sound recognition for children given their simplicity (Figure 13).

---

[43] http://www.bbc.co.uk/worldservice/learningenglish/grammar/pron/
[44] By using X-rays, for example.

They claim that three-dimensional articulatory models may pose extra processing difficulties for certain audiences (like children) and argue in favour of more easily understandable 2D models. In their view, two-dimensional views include all the essential information of speech movements and are less complex. This point is also raised by Massaro et al. (2008), who claim that the external view of the face seems to be processed more easily than an internal view showing the tongue, palate and velum. They point out that this may depend on the sound being targeted, claiming that perhaps talking heads should modify the view they present depending on the phoneme they are illustrating. For example, while the articulation of /p/ may not require all the information offered by a 3D view, 3D may be very useful in order to illustrate the lateral pronunciation of /l/.

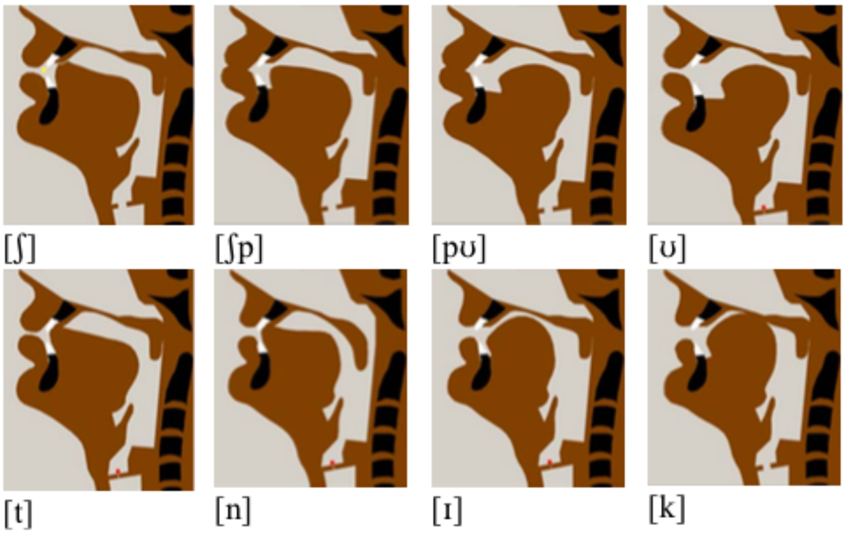


Figure 13. Example of the 2D animations used by Kröger et al. (2008) showing sound articulation from a midsagittal view. Taken from Kröger et al. (2008, p. 2639)

Additionally, the quality of graphics is something to consider when devising this type of software. Versions of these talking heads are starting to appear in apps for smartphones and tablets (see Massaro et al., 2012; Psyentific Mind, 2012),[45] but in some cases, even the computer version does not look very natural despite using state-of-the-art technology (see e.g. Figure 11 for the talking head employed by Engwall, 2008, 2012; or Figure 14 for the one used by Massaro et al., 2012). If high-quality resolution is not available or poses constraints on programming the software, perhaps 2D animations showing place of articulation would suffice, as suggested by Kröger et al. (2008). As a case in point, the *Sounds of Speech* app, developed by the University of Iowa Research Foundation (UIRF,

[45] Link to iBaldi (Psyentific Mind, 2012): https://itunes.apple.com/es/app/ibaldi/id504464546?mt=8

2014), offers 2D animations that can be considered quite informative, at least for some phonemes (see Figure 15).[46]

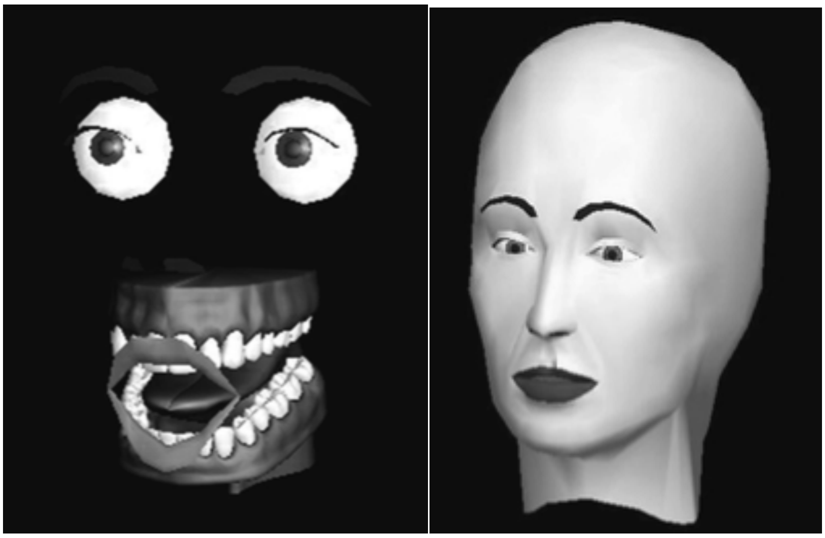


Figure 14. Example of the talking head employed by Massaro et al. (2012) showing an internal view of articulations (on the left) and a front view of their canonical head (on the right). Taken from Massaro et al. (2012, pp. 320 and 330)

Research should investigate whether students make progress when working on their own with these apps. So far, research has focused on extremely controlled laboratory environments and training has been sometimes limited to very short periods of time and to very small numbers of participants. Future studies should also address whether working on the app yields the same results as working in a computer, as the app allows users to practise in contexts that may not be as quiet as one's home (bus, train, street). In addition, research should address the teaching of other aspects of pronunciation. So far, investigations have mostly focused on the training of consonants, for which these representations are undoubtedly promising given that they show place of articulation very clearly. Nonetheless, it would be interesting to see how useful these agents are in order to teach vowels (not having such clear representations of articulation) or aspects of connected speech, such as linking (if using articulatory information), or intonation (if showing a frontal view of the face, by conveying attitude with facial expression and intonation, for example).

[46] Link to *Sounds of Speech* (UIRF, 2014): https://itunes.apple.com/us/app/sounds-of-speech/id780656219?mt=8

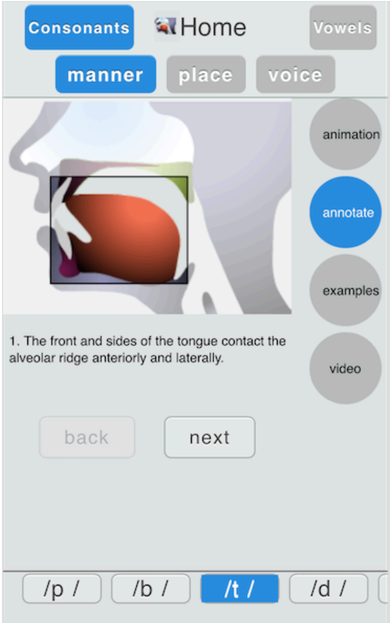


Figure 15. Example of 2D animation offered by the *Sounds of Speech* app (UIRF, 2014)

Finally, it is important to take into consideration that the potential of animated agents may also be determined by the learners' first language. As Hazan et al. (2005) indicate, languages like Chinese or Japanese have a very low visual influence as compared to other languages. Thus, learners from these L1 backgrounds may require additional training in order to be able to attend to the articulatory information available through the visual channel.

b) Productive enhancement

As pointed out in the section dealing with visual enhancement, the techniques adopted for the productive enhancement of pronunciation are closely linked to those for enhancing perception. Most of the tools described in the above sections could also be considered to facilitate production, as they are used as a means to provide learners with feedback on their productions, rather than as mere perceptual enhancement. Were it not for this feedback, most learners would not repeat most of the utterances they mispronounce, as they would not be in a position to judge whether their pronunciation is adequate or not (see e.g. de Bot, 1983; Hew & Ohki, 2004). Nevertheless, they were included under perceptual enhancement because it is the prior perception of problematic aspects in their output that prompts them to make further attempts in their production.

In light of the inconveniences of using visual representations of speech for autonomous practice reviewed above, Automatic Speech Recognition (ASR) provides a promising solution to help learners practise on their own without requiring any knowledge of phonetics.

Automatic Listeners: Automatic Speech Recognition

Speech recognition[47] is defined as 'the process of enabling a computer to identify and respond to the sounds produced in human speech' (Apple Inc., 2014). This can range from merely recognising the words a person utters and transcribing them, like the traditional Microsoft Word dictation function or speech-to-text applications such as Dragon Dictation, to recognising what someone says and reacting to it, such as iPhone's Siri, which allows users to give voice commands such as creating, modifying or deleting a reminder, setting an alarm, dictating a message/email and sending it, or even having limited conversations with the phone. This technology is increasingly becoming common in our daily lives, for example in automatic answering machines that transfer you to one department or another depending on what you say, or in various accessibility tools for people with special needs.

Despite the central importance of perception in the acquisition of pronunciation, a great amount of productive rehearsal is also required in order to automatise the articulatory patterns of the foreign language. Thus, given that one of the greatest advantages offered by technology is autonomous practice, ASR technology opens up an enormous range of possibilities for CAPT, as it provides a way of evaluating learners' output objectively, immediately and automatically. As Hincks (2003, 2005) points out, ASR has long held the exciting promise of allowing learners to have a truly communicative, feedback-providing conversation with a computer. This technology has been used in a number of commercial applications (e.g. talking heads, courseware like Tell Me More, Talk To Me, etc.) that allow learners to practise outside the classroom and be automatically evaluated by the machine.

---

[47] Some authors consider technologies such as feedback through speech visualisation as examples of Speech Recognition because the computer somehow processes speech from the student (e.g. Aist, 1999). However, it could be argued that in spectrograms, waveforms and the like, the machine does not really 'recognise' speech, but merely offers visual representations of the acoustic characteristics of the speech signal. Thus, in this dissertation, ASR will only be used to refer to those tools where the machine is supposed to react to learners' productions by identifying what they actually say (e.g. by transcribing speech into text, by responding in a dialogue depending on the learners' answer, or spotting learners' mistakes).
In addition, a distinction could be made among Automatic Speech Recognition, Automatic Pronunciation Assessment, and Automatic Error Detection. The first could be understood as a system that merely 'recognises' or 'understands' what a user says; the second provides a score on learners' pronunciation, without necessarily indicating where mistakes are or how to correct them; and the third should be able to spot mispronunciations and hopefully offer feedback on how to correct them (see e.g. Eskenazi, 2009). Nonetheless, for the sake of simplicity, ASR is used indistinctively in this dissertation to refer to the three types, as any system that is capable of 'detecting' errors must first 'recognise' what the user says.

ASR allows the provision of feedback on persistent production errors in a non-threatening context (Derwing et al., 2000), it does not rely on the learners' perceptions (Ehsani & Knodt, 1998) and it can show learners' performance over time by keeping a record of recurrent mistakes and improvements (Herron et al., 1999). This is highly advantageous as compared to courseware that encourages learners to record their pronunciations and compare them to a model stored, given that students are not in the position to judge their own productions because of their 'phonological deafness' (Renard, 1979), as a result of their L1 (i.e. their inability to perceive the FL phonology adequately, and consequently, their mispronunciations).

Pronunciation tutors using ASR technology can be employed to prompt students to repeat words or phrases that have been mispronounced, with the advantage that it frees teachers from this time-consuming activity and offers a degree of consistency that is not always easy among humans (Hincks, 2003). Besides, it seems particularly suitable for correcting reading aloud practice, which fosters literacy skills and helps learners establish the associations between sounds and orthography (Ehsani & Knodt, 1998) – something particularly challenging for certain EFL learner groups (see e.g. Mompean & Fouz-González, in press; Monroy, 2001).

As far as CAPT is concerned, ASR has mainly been used for three purposes. First, to convert speech into text, illustrating clearly what the machine understands and what it does not (e.g. Coniam, 1999; Fouz-González, 2012). Second, to react to what users say in a simulated conversation and continue the conversational path depending on what the user decides/says – an indirect measure of learners' intelligibility, for example in virtual worlds (e.g. Morton & Jack, 2005; Morton, Gunson, and Jack, 2012; or courseware like Tell Me More or Talk To Me). And third, as a means of pronunciation scoring, which can be used for language testing (Hincks, 2001), or to offer learners feedback on their pronunciation, for example in talking heads (Engwall, 2008), courseware like Tell Me More (Elimat & Abuseileek, 2014), or dialogue systems for specific purposes (Walker, Trofimovich, Cedergren, & Gatbonton, 2011).

Requirements & some findings

An essential condition ASR systems intended for use with non-native speakers should meet is to understand the speech of foreign language learners. The fact that an automatic machine understands a native speaker when phoning a call centre does not mean that the same holds for non-native speakers. As noted by Benzeghiba et al. (2007), research corpora and speech models employed for comparison in ASR do not commonly comprise samples representing the significant variation that occurs in speech, such as children or elderly speech, or that of foreign language learners, types considered to differ significantly from the 'standard form' (p. 765). This poses considerable demands for ASR, as the system must be able to recognise samples of non-native language from a range of different accents as well as pronunciations including subtle deviations from standard native pronunciations (Ehsani & Knodt, 1998), both at the phonetic and grammatical levels (Godwin-Jones, 2009).

Focusing on ASR applications for CAPT, that is, for the specific purpose of pronunciation training, the ultimate goal should be that non-native speech can be assessed in the same way as a human expert would (Eskenazi, 2009). ASR applications should be able to detect those problems that impose difficulties in intelligibility for a human listener (Derwing et al., 2000). However, as shown by several studies, the accuracy of these systems decreases significantly with non-native speech (e.g. Coniam, 1999; Derwing et al., 2000) and ASR ratings do not always offer the expected ratings as compared to those given by humans (Derwing et al., 2000; Kim, 2006). In spite of advances in the field, an acceptable level of reliability is only guaranteed when the tasks are simple and the set of utterances from which students can select a desired response is kept to a restricted set (Menzel et al. 2001), something that limits the usability of this technology for spontaneous practice. As a case in point, while the software tested by Derwing and colleagues reached the advertised accuracy with native speech (90% accuracy), it decreased to 71-73% with non-native speech despite the fact that subjects were highly proficient in English. The authors claim that 'their speech was only slightly less intelligible to human listeners than native English speech' (Derwing et al., 2000, p. 601), something they identify as the target endpoint for many L2 learners.

An ideal ASR system for pronunciation training should recognise everything the user says, point out those areas that are most problematic depending on the user's priorities (be it intelligibility, comprehensibility or accuracy), and then offer explicit feedback on how to improve. In this regard, Eskenazi (2009) identifies two common approaches to evaluate learner pronunciation commonly employed in ASR: global pronunciation scoring and individual error detection. The former provides a general pronunciation assessment on the overall impression of the user's speech[48] while the latter addresses individual aspects of pronunciation, such as segments, or suprasegmental phenomena.

Despite the fact that global pronunciation scores are helpful in order to offer a general indication of a learner's proficiency, they are not as informative as learners need, given that they do not provide specific information on the nature of their mistakes (Strik, Truong, de Wet, & Cucchiarini, 2009). Following Levis' (2007) example, telling students that their pronunciation of an utterance is of 62% is only as useful as saying 'this means that you often have words that are mispronounced' (p. 192). Learners are neither told what their specific mistakes are, nor how to correct them. Thus, besides receiving a score, learners should always know what their mistakes are and be offered instructions on how to improve it (Menzel et al., 2001; Neri et al., 2002). Menzel et al. (2001) claim that even word-level scoring that permits learners to identify the mispronounced word is not enough. They maintain that it is only at the phone level that the system is in a position to offer detailed explanations concerning the nature of the mistake and provide instructions for improvement. Hence, in most cases, the training offered by ASR systems is ultimately limited to trial and error (Engwal & Bälter, 2007).

With regard to individual error detection, one of the most ambitious projects exploring the use of ASR for FL pronunciation training is the ISLE (Interactive Spoken Language Education) project (Atwell, Herron, Howarth, Morton, & Wick,1999; Menzel et al, 2001). Its main objective was to stand out among other ASR technology by offering high reliability when localising mispronunciations in specific phones and by offering constructive feedback to the students. That is, not only mentioning that an error has occurred, but also indicating what learners have to do to correct it (Herron et al., 1999). An additional advantage of the system was that it was devised to understand foreign-accented speech, as opposed to previous adaptations of ASR software designed for NSs and then reutilised for NNSs. The system was initially designed to understand Italian and German learners, but the authors claim that the underlying mechanisms could be extrapolated to other languages. Nevertheless, the levels of performance were far from satisfactory, as the system yielded a high number of false alarms (i.e. adequate pronunciations that were scored

[48] Because the goal of this section is to describe the possibilities this technology offers for pronunciation teaching, detailed explanations of how this technology works are not included here. The reader is referred to Huang, Acero, and Hon (2004) or Witt (1999) for a comprehensive description of these systems and how they work, or to Eskenazi (2009) for a review.

as incorrect) and the number of correct detections was rather low both in terms of phonemes and lexical stress (see Menzel et al., 2001).

Research investigating the potential of ASR systems for pronunciation training has offered encouraging results, although this technology needs to improve substantially before learners can use these systems autonomously and rely entirely on their judgements. For example, Neri, Cucchiarini, and Strik (2008) tested the effectiveness of feedback offered by an ASR system in order to improve learners' pronunciation of a number of problematic Dutch phonemes. They compared a group receiving CAPT training with feedback offered by ASR, a group also training with CAPT but without feedback, and a control group receiving no training. Even though the differences among the three groups were not significant, the group receiving ASR feedback showed considerable improvements as compared to the group without feedback. In a similar line, Liakin, Cardoso and Liakina (2014) explored the effectiveness of ASR feedback on the pronunciation of French /y/ by anglophone speakers in Montreal. They also compared three groups of learners: one receiving ASR feedback on their mobile devices, another receiving the same pronunciation activities but receiving feedback from a teacher, and a control group that did not receive feedback on their pronunciation. Their results support the effectiveness of the ASR system, as the group receiving ASR feedback was the one that obtained significant improvements in the production of /y/.

Although using a rather small group of students (five in the treatment group and one in the control group), Burleston (2014) found that training with an ASR system helped Mandarin learners improve considerably their intelligibility of six English sound contrasts and generalise improvements to untrained words. It is important to mention that the task was kept quite simple for the recogniser, as words were practised in isolation and the machine knew what to expect (i.e. it was not spontaneous production); the ASR might not have worked so well for communicative activities. Nonetheless, the technology is undoubtedly promising for practising those aspects learners find most problematic in a controlled manner.

Elimat and AbuSeileek (2014) compared the effectiveness of the Tell Me More courseware as used by three groups of students receiving instruction individually, in pairs, and in groups, with a control group following traditional listen-and-repeat pronunciation instruction. Even though they report that their treatment groups obtained significant gains as compared to the control group, their results must be taken with caution, given that their description of the data treatment is somewhat vague and the premise of the study methodologically flawed. They claim to test the effectiveness of ASR to improve 'the students' performance in pronunciation' (p. 30), but the pre- and pos-test design combines an evaluation of both perception and production all at once, including activities in which learners discriminate auditorily among three minimal pairs, gapped dialogues that evaluate word identification, as well as different types of productive tests. Moreover, for the interpretation of data they simply provide the mean performance for all the subjects in each group for all the activities (mixing perception and production). It might be the case that participants had improved considerably in terms of discrimination skills for the minimal pairs presented (and obtained a 'good average performance'), but they had not improved one single aspect of production. Therefore, these findings cannot be interpreted as an example of ASR helping learners improve their production.

A final example is the study by Mayfield Tomokiyo, Wang and Eskenazi (2000). They used the Fluency system to train the pronunciation of /θ/ and /ð/ with speakers of different L1s. Even though there were no significant differences between the experimental group and a group receiving classroom instruction, the group receiving ASR training managed to reduce the error rate substantially (47.2%). It is important to mention though, that learners in this study received corrective feedback in the form of articulatory information provided by a side headcut and a front view of a model's face. This should not be taken as a limitation. In fact, this is the approach proposed by Engwall (2008) as a possible solution to the problem of unreliable ASR feedback; as recommended by Neri et al. (2002), feedback should not only pinpoint problematic areas, but also indicate learners how to improve. However, the effectiveness of the technology in this particular study cannot be ascribed to the ASR detection alone, but also to the visual feedback offered, commonly included in this type of courseware.

Limitations

In spite of the enormous potential of ASR for pronunciation training, there are a number of limitations that must be acknowledged. As Eskenazi (2009) notes, deficiencies in ASR technology have commonly led researchers to focus on its strengths and avoid encountering its limitations, shaping users' expectations to match the capabilities of the technology rather than adapting the technology to users' needs.

Despite the fact that ASR has come a long way from early forms of voice recognition technology requiring every word to be pronounced in its citation form and even emphasise pauses between words (Coniam, 1999), state-of-the-art recognition does not yet allow to have a genuinely spontaneous conversation with a machine. Researchers have often ascribed this limitation to the ASR technology (Kenning & Kenning, 1990; O'Brien, 2006). However, given the high reliability obtained in terms of recognition by modern ASR systems such as Siri,[49] perhaps what makes extemporaneous conversations impossible is not the entirely the recognition module, but the underlying artificial intelligence, as the machine does not only need to understand users, but also have an answer to everything they say. In any case, the solution that is commonly adopted consists in restricting the language users are expected to produce to a 'manageable subset', something that is considered to be 'obviously restrictive and quite unlike natural communication' (Kenning & Kenning, 1990, p. 24).

Although having spontaneous conversations with a machine is indisputably auspicious for FL learning, the real promise for CAPT is the automatic detection of learners' mistakes and the possibility to offer feedback on their pronunciation. In this regard, although ASR can *understand* what users say to a considerable extent (see e.g., Derwing et al., 2000; or Siri), current technology is not yet able to adequately identify all the mistakes made by foreign language learners (Menzel et al., 2001; Neri et al., 2008). One of the major limitations is the provision of erroneous feedback (Neri et al., 2002), either in the form of false positives (i.e. telling learners that something is correct when it is not), or false negatives (i.e. telling learners that they pronounced something wrong when they did not – Eskenazi, 2009).

Even though some (Menzel et al., 2000) maintain that students could benefit from merely directing their attention to the problems spotted by the system, the experience may be quite frustrating for users if mistakes are not detected or are detected incorrectly (Derwing et al., 2000), as the machine is supposed to be some kind of 'expert' they can rely on. As Wik and Hjalmarsson (2009) point out, erroneous feedback will have a much stronger negative impact on foreign language learners than if it were native speakers who were misunderstood by an ASR system; the former will probably assume that their language proficiency is low and ascribe the score to their performance, while the latter will be fully aware that they have not done anything wrong. As Neri et al. (2002, p. 459) write '[r]eliability is crucial in language learning: nothing could be more confusing for a learner

[49] Siri recognises quite a large percentage of what users say, even with background noise. It also has an answer to *almost* everything users say. It is well-known for her funny responses to questions such as 'do you love me?' or 'what is the meaning of life?'. For examples visit: http://www.macworld.co.uk/feature/iosapps/funny-things-ask-siri-3412780/

than a system reacting in different ways to successive realizations of the same mistake'. Once learners suspect the system is not reliable, they will lose all their confidence in the system, therefore rendering the machine useless – as it would be impossible to know when the machine is scoring well and when it is not.

In light of the above, and in line with Eskenazi's (2009) claim that it is often users (be it teachers or students) who need to adapt to technology, a common 'solution' is to simplify the type of feedback offered. In the same way that detection is sometimes simplified by reducing the number of possible answers from users (i.e. by limiting the input that is expected from them to a few options), the assumption with feedback is that by making half the number of decisions, the machine will make half the mistakes (Neri et al., 2002). That is, merely informing learners that a word has been mispronounced is easier than locating the specific problematic phoneme (Levis, 2007), and consequently, easier than offering them instructions on how to correct it. Nonetheless, despite the fact that the majority of recommendations on how to overcome the obstacles in ASR are quite sensible, some of the advocated 'adaptations' have dubious pedagogical value. For example, some of the recommendations given by Menzel et al. (2001) include making activities 'difficult enough for the target user group in order to produce enough pronunciation errors' or targeting those 'errors that can be diagnosed with higher degrees of reliability' (p. 74). These suggestions seem to lose track from the nature of pronunciation teaching. On the one hand, they encourage the use of activities that are difficult per se, as though forcing learners to make mistakes was something positive for their learning. On the other hand, learners would be bound to practise exclusively with the aspects the machine is able to detect and correct (sounds, stress, etc.). Moreover, this idea clearly runs contrary to the communicative function of pronunciation, as the conversational flow would be interrupted constantly and learners will probably become discouraged by a 'tutor' that only trips them up because it is the only thing it can do.

As noted above, given that the main problem with ASR systems used for CAPT is feedback (Neri et al., 2002), a possible alternative is to illustrate how to correct mispronunciations with visual feedback offered by animated agents (Engwall & Bälter, 2007). Apart from the numerous benefits of these agents presented above (motivation, ease of interpretability of articulatory gestures, users' willingness to spend time with the system due to their human-like appearance, etc.), one of the clearest advantages of talking faces using ASR is that the machine does not need to understand the nature of the mispronunciation in order to offer feedback. By simply detecting an error, the system could offer articulatory illustrations on how to pronounce a particular segment, or if impossible to detect errors to that level, illustrate the pronunciation of the whole word, perhaps allowing learners to visualise segment by segment (therefore also allowing them to spot what they may be doing differently).

Besides, as noted above, these tutors could offer feedback in a manner that is similar to that of a human teacher, not only in terms of the type of feedback offered, but also in the way that a human would offer it (see Engwall, 2008; Engwall & Bälter, 2007). An innovative approach that should be mentioned again here is the dual type of training proposed by Wik and Hjalmarsson (2009), where two different agents are employed for different purposes. The user could maintain a conversation with an animated agent (DEAL) and be directed to a different agent (Ville) for pronunciation reinforcement if errors are detected.

As Levis (2007, p. 192) notes, '[t]he failure of ASR systems (developed for native speakers) to successfully handle nonnative speech does not mean that ASR is useless in pronunciation training. Rather, it means that new ASR systems with goals more appropriate to language learning contexts are needed'. However, there will always be idiosyncratic errors that machines will be unable to detect (Neri et al., 2002). Furthermore, even if this technology were perfect in terms of recognition and diagnosis of errors, it would encounter obstacles that go beyond the potential technological limitations. As Pennington and Esling (1996) note, 'there is no definite quantifiable standard as to how far a learner's pronunciation might deviate from the model and still be acceptable' (p. 173). This argument is echoed by Witt (1999), who claims that there is not such thing as an absolutely 'correct' pronunciation and points to the existing variation across accents and varieties of English. If a system is trained to understand British or American English, for example, what varieties within those accents should the system recognise as good? It would be preposterous to judge the pronunciation of *must* as [mʊst] as incorrect because it does not belong to the standard British English (Southern variety) with /ʌ/. A possible solution would be that users could choose the variety they want to practise, but as Hincks (2003) has noted, because building ASR systems is costly and time-consuming, it would be unfeasible to train systems to understand certain languages or accents.

Even though ASR is not reliable for genuine communicative practice, it holds great promise for autonomous practice with aspects of the FL phonology learners need to automatise. Although drilling is commonly criticised when it implies decontextualised practice devoid of meaning (see e.g. Baker, 1992), drilling useful sentences in communicative activities can be quite useful in order to automatise articulatory habits, especially because some articulatory gestures may require a re-establishment of fossilised muscular movements after years of repetition in the mother tongue. Drills are not necessarily mechanical or incompatible with communicative activities (see Fraser, 2001; Isaacs, 2009; Paulston, 1972). Drilling should not be conceived as something limited to minimal sound pairs. As a case in point, communicative dialogues could be used in order to help learners practice sentence stress in English, which, at the same time could help learners practise vowel reduction. Learners could be given a range of options containing common problematic aspects for their L1 group and then offered controlled practice with aspects they need to automatise. If the machine knows what to expect, the system would only need to judge how well the user pronounces certain sound, or the intonation used for a particular communicative function, for example. As mentioned in the sections above, learners could work on the automatisation of articulatory habits of the FL phonology autonomously and then put this into practice in class with other classmates in real conversations.

In the absence of adequate ASR feedback, the use of speech-to-text applications that transcribe what the user says is a possible alternative (Coniam, 1999; Fouz-González, 2012). This would allow a clear representation of the user's intelligibility in extemporaneous speech, as the student could read what the machine understood (in orthography), being also able to identify the exact location of mistakes. For instance, if a student read the sentence *Who are you going to boat for?*, (s)he would immediately realise that there was a problem with the pronunciation of /v/. Since the machine would offer a script on what the learner said, (s)he could then go back to the script and practise with those aspects the machine did not understand in a more controlled environment. Once the learner became aware of the nature of mispronunciations (in this example, the /b-v/ pair), (s)he could be offered specific activities targeting that pair, and in that case, the ASR system could offer a score on how well the speaker pronounces that particular target item (or articulatory information on how to pronounce it, if using talking faces).

O'Brien (2006) states that 'the pedagogical effectiveness of courseware that employs ASR technology is greater than that of courseware that requires learner self-evaluation of pronunciation' (p. 133) and defends its usefulness in certain controlled contexts. Even though ASR can be useful for certain types of pronunciation training, it is important to take into account that learners who are 'outperformed' by machines in terms of pronunciation evaluations have commonly received no explicit training in pronunciation.

For example, in studies showing the advantages of visual feedback over auditory feedback alone (e.g. Akahane-Yamada et al., 1998; de Bot, 1983; Hardison, 2003; Hew & Ohki, 2001, 2004; Liu et al., 2007), learners in the 'audio alone' groups had to rely on their own intuitions about pronunciations. Learners in those control groups are commonly ignorant of how pronunciation works, as they are often not given any explicit pronunciation training, and consequently, it is not surprising that *any* machine will be able to evaluate their mistakes better than themselves. As mentioned above, FL learners are said to suffer some kind of 'phonological deafness' (Renard, 1979) that prevents them from perceiving the FL phonology adequately as a result of the phonological 'sieve' (Trubetzkoy, 1939) of their mother tongue. Thus, expecting linguistically-naïve learners to rely on their own intuitions about their pronunciation would be like asking someone who has never studied French to evaluate the grammar of his peers speaking French.

In light of the above, we can conclude that foreign language syllabi should never be restricted to 'what a machine can do'. Teachers should design a pronunciation syllabus tailored to their learners' needs (paying special attention to those aspects that are most problematic for their L1 group) and then select the technology that best satisfies their needs. Given that none of the above-reviewed technologies is perfect, technology should never be understood as 'the holy grail', but as a supplement to classroom instruction that allows more individualised attention, extra-time, a private environment, and as something which can besides enhance the way particularly problematic aspects are presented.

1.3.3 The CALL we miss: Using accessible and affordable technologies to train FL pronunciation

The above sections have reviewed a number of studies providing ample evidence that pronunciation *can* be trained and that directing learners' attention to particular pronunciation points can help them improve aspects that would otherwise be extremely difficult to modify. Nevertheless, despite the great promise the above technologies offer for pronunciation training, they do not seem to be entirely suitable for autonomous practice yet.

Despite the numerous advantages CALL offers, one of the most-cited claims in the literature is the mismatch between technology and pedagogy (Kenning, 2007; Levi, 1997b; Neri et al., 2002; Pennington, 1999). A great number of CALL and CAPT applications seem to be 'the result of a technology push, rather than a demand pull' (Neri et al., 2002, p. 442), commonly using flashy features to impress users despite their limited usefulness for language acquisition. While 'the potential pedagogical effect of the technological tools used in L2 instruction […] is inherently dependent on the particular theoretical or methodological approach that guides its application' (Salaberry, 1996, p. 7), research has focused excessively on the potential benefits of technological advances instead of exploring the way technology can be used to support language teaching and learning (Liu, Moore, Graham, & Lee, 2003).

As a case in point, one of the areas that has received most attention in CAPT research is ASR feedback, regardless of its limited reliability when used with foreign language learners. Although research is unquestionably needed in order to make progress in any area, the fact some commercial applications implement technology with no regard for their quality (e.g. the aforementioned examples of the speech synthesis – */aɪ el el gəʊ/ for *I'll go*) suggests that developers continue to guide their decisions by the capabilities of technology rather than adapting technology to the needs of the user.

Furthermore, as Nelson and Oliver (1999, p. 101) recommend, CALL should not merely 'replicate textbooks and workbooks […]. We should not be using a computer just because we can and we should not be using more advanced (i.e. complicated) technologies when something simpler will do'. Technology should be used in order to truly enhance learning, using activities that genuinely exploit the capabilities of the technology being used, not simply 'rehashing' old techniques and doing with technology the same we could do without it (Colpaert, 2004; Setter & Jenkins, 2005).

CALL materials do not necessarily need to be used as stand-alone products, but can also work perfectly when integrated into other activities (Hardisty & Windeatt, 1989). In fact, another common claim in the literature is that CAPT studies are often conducted under extremely laboratory-like settings (Olson, 2014; Wang & Munro, 2004), which may not be very representative of what may happen in a real classroom. Thus, research exploring the effectiveness of different CAPT applications in real classroom settings is in order.

Technology can enhance pronunciation training in ways a teacher cannot and that were unthinkable before, such as offering different exemplars of the same sound on demand (with different voices – male and female – and in different accents), illustrating internal views of articulatory movements, or making aspects such as intonation more 'tangible'. However, as illustrated in the previous sections, many of the technologies employed in CAPT require the pedagogical intervention of an expert (e.g. spectrograms or waveform representations of the learners' speech), and when devised with autonomous practice in mind (such as talking heads, or courseware using automatic speech recognition), they have proven to be unreliable when dealing with non-native speech.

Thus, since 'the perfect tool' does not exist and is not likely to materialise in the near future, research should concentrate on finding alternative ways in which technology can enhance and support pronunciation instruction. Given that the automatic evaluation of foreign speech is not entirely reliable, the approach adopted here is that of offering learners pronunciation training that empowers them to work on their pronunciation autonomously. The remainder of this section will briefly outline the three technologies explored in this dissertation, namely podcasts, social networking services (in particular, Twitter), and smartphone applications.

These three technologies are considered to be accessible and affordable technologies that are easily interpretable to learners. They are either free or relatively cheap, and some of them are already used by students as part of their e-routine.[50] This should facilitate their implementation as part of classroom instruction or as a complement to it, extending the physical and time boundaries of school.

Podcasts

Podcasts are defined in the Oxford Dicitonary of English (Apple Inc., 2014) as digital audio files 'made available on the Internet for downloading to a computer or portable media player, typically available as a series, new instalments of which can be received by subscribers automatically'.[51] The word *podcast* comes from the first Apple's audio and video player (the iPod), and the word *broadcast* (Sharma & Barret, 2007), although today they can be played in almost any device (computer, tablet, smartphone, or mp3 player).

Podcasts have been compared to radio or TV shows in that they are often released periodically (Dudeney & Hockly, 2007). They are usually updated on a regular basis (daily, weekly, etc.), allowing users to subscribe to the podcasts they like so that the new episodes are automatically downloaded to their computer, tablet, or smartphone as they are released. They offer an incredible range of topics (food, comedy, history, news, etc.) with podcasts

[50] The term e-routine is used here to refer to the students' daily use of electronic devices (e.g. smartphone apps such as Whatsapp) or social media such as Twitter, Facebook, etc.

[51] It should be noted that podcasts also exist in video format (often referred to as vodcasts). Nevertheless, since the specifics of the technology have been described in detail by numerous authors, for the purposes of this dissertation, only the essential characteristics of the technology will be dealt with here (for a more comprehensive description of this technology see e.g. Godwin-Jones, 2005a; McBride, 2009a; Rosell-Aguilar, 2007, 2009; Stanley, 2006).

of varying lengths (from three minutes to more than an hour). Podcast creators often make their podcasts available from their website, although the easiest and most comfortable way of finding and managing them is through iTunes.[52]

Several authors note that listening to audio on the Internet is nothing new, highlighting the possibility to subscribe and automatically download podcasts as their defining feature (Stanley, 2006; Rosell-Aguilar, 2007). However, a broader meaning of the term podcast can be adopted in order to refer to the majority of 'downloadable sound files on the Internet, without them needing to be either syndicated,[53] nor one in a searies of episodes', which covers a much wider range of resources (McBride, 2009a, p. 154). In fact, one of the reasons that make podcasts extremely appealing for language learning, and especially for pronunciation practice, is that they are not exclusive to professionals; anyone can easily and freely create their own podcasts and distribute them through the Internet. Hence, despite the fact that podcasts employed in Studies 1 and 2 in this dissertation were indeed part of a series (BBC 6-minute English – see below), the term podcast is also used to refer to the audio files students recorded.

The advantages this technology offers for FL learning seem obvious, as they provide learners with a virtualy unlimited source of input – one of the core elements in language acquisition (Krashen, 1982, 1985); but now also offer users the possibility to record their own podcasts and therefore produce output, another essential element in SLA (Swain, 1985, 2005). The fact that podcasts are extremely portable and exist in a variety of topics and lengths allows learners to select from a wide range of materials according to their needs and interests and listen to them anywhere and at any time.

Podcasts have been used to provide speech language therapy students with practise on phonetic transcription (Knight, 2010), to help FL learners work on their listening skills (Kim, 2013; O'Bryan & Hegelheimer, 2007; Weinberg, Knoerr, & Vandergrift, 2011), foster and analyse learners' oral productions – including pronunciation – (Tomé, 2010), or help learners work on specific aspects of their pronunciation (Ducate & Lomicka, 2009; Mbah, 2014; Lord, 2008). However, the majority of studies in the literature either simply

---

[52] It is Apple's software for searching, downloading, managing and playing all the digital content that can be transferred to Apple products – iPod, iPhone, iPad, etc. However, it runs on every computer and users do not need to have an Apple device to play podcasts. They can be played on any MP3 player, computer or similar device.

[53] Syndicated here comes from RSS (Really Simple Syndication) feeds, 'a standardized system for the distribution of content from an online publisher to Internet users' (Apple Inc., 2014). That is, a file that is published on the internet and to which users can be subscribed.

describe the technology and/or its possibilities, or assess the learners' self-perceptions of learning or their attitudes towards podcast implementation. To date, very few studies assess the potential of this technology empirically.[54]

The study by Lord (2008) investigated podcasts' potential to improve the learners' general pronunciation ability, their attitudes towards Spanish pronunciation as well as their perceptions regarding the use of this technology in the project. Lord did not contemplate podcasts as a source of input provided by a native model; she concentrated on the possibities offered by podcasts in order to foster learners' output and self-monitoring skills. She used a group of 16 students enrolled in a Spanish phonetics class. Participants were required to record a total of six podcasts that covered different problematic aspects of Spanish pronunciation, namely, a number of consonant and vowel segments and linking. The podcasts learners had to record were mostly based on written texts and tongue twisters, although one of the podcasts offered a more spontaneous type of production in which learners talked about their language background and their experience learning Spanish. Students were required to upload their podcasts onto a shared site and evaluate each other's podcasts leaving constructive comments on how to improve their pronunciation. Lord's data reveal a positive effect of training on the pronunciation of most participants. Furthermore, learners' attitudes towards pronunciation improved after training and they claimed to have enjoyed the project and considered it beneficial.

In a similar vein, Ducate and Lomicka (2009) examined podcasts' potential to improve accentedness and comprehensibility of 22 students in German and Italian courses as well as their attitude towards pronunciation. They were required to record a total of eight podcasts. Five of them were scripted and the other three were extemporaneous accounts of their experience learning those two languages. In this study, the three scripted podcasts also included model podcasts as input that exemplified learners how to pronounce their subsequent recordings. As in Lord's (2008) study, participants were also required to upload podcasts to a blog and comment on their peers' podcasts. Ducate and Lomicka's (2009) results show that despite slight improvements in some of the tasks, the learners' accentedness and comprehensibility did not improve over the course of training, nor did the learners' attitudes towards pronunciation. Given that other studies have

---

[54] The study by Mbah (2014) also presents the results of an empirical study with a pre- and post-test evaluation of the learners' pronunciation. Nevertheless, the procedure employed both for testing and training – including stimuli – is almost entirely omitted in the paper. Given the impossibility to assess the validity of their findings, the results are not further discussed here.

fostered improvements in pronunciation in much shorter periods of time (e.g. Lord, 2008; Saito, 2013), they point to a possible lack of focus on form and explicit instruction as one of the reasons why learners did not improve.

Finally, Tanner and Landon (2009) conduced a study aimed at evaluating the effectiveness of self-directed oral readings to improve learners' perception and production of pausing, word stress and sentence-final intonation, as well as comprehensibility.[55] Their oral readings consisted in raising the learners' awareness of the above-mentioned features by listening to short passages read by native speakers and marking the location of the target features in the passage, finally practising reading the passage themselves. Participants in their study improved their perception of pausing and word stress considerably, as well as their controlled production of the latter. Nonetheless, no improvements were found in the learners' comprehensibility.

Smartphone apps

Apps are defined as 'self-contained program[s] or piece[s] of software designed to fulfil a particular purpose, […] especially as downloaded by a user to a mobile device' (Apple Inc., 2014). Smartphones are incredibly versatile thanks to the number of apps at their disposal that add new functions to the phone. Today's smartphones are a sort of Swiss-army-knife enabling users to use their phones as dictionaries, torches, to control their diet or even track their runs via GPS. As Godwin-Jones (2011, p. 2) puts it, with today's smartphones 'what used to be phones with added-on computing capabilities have morphed into mini-computers which can also make pone calls'.

Even though the three technologies explored in this dissertation could be considered to fall within the field of mobile learning, in that they can all be used in mobile devices, podcasts and Twitter are not restricted to be used in portable devices, while apps usually are.[56] This offers advantages such as being able to access information and use the app whenever and wherever users need, but also limitations such as screen size or potentially noisy environments that may not be very appropriate for activities that demand concentration (see Stockwell, 2010).

---

[55] Even though they did not use podcasts as such, they are also included in this section given that the procedure is very similar to the approach adopted in this dissertation with podcasts and to the input-output sequence followed in Ducate and Lomicka's (2009) study. Participants in Tanner and Landon's (2009) study received input in Mp3 format and recorded their output with Windows voice recorder.

[56] As stated in the definition, apps are often originally devised for mobile devices (smartphones, tablets, smart watches, etc.). However, Apple desktop computers and the new versions of Windows are also giving their software certain app-like appearance, both in design and in the way they are purchased and downloaded.

Research has explored the potential of mobile devices for different language learning purposes, such as mobile phones to teach vocabulary through SMS (Levy & Kennedy, 2005; Kennedy & Levy, 2008) or e-mail (Thornton & Houser, 2005), portable consoles to work on listening and reading skills (Kondo et al., 2012), FaceTime conversations through iPads to foster oral skills practice (Lys, 2013), or multimedia messages to improve pronunciation (Saran, Seferoglu, & Cagiltay, 2009), to name a few. However, despite the fact that some of these studies have used apps as part of their training procedure (e.g. Kim, 2013; Lys, 2013), few studies have focused on the actual potential of particular apps to train FL aspects (e.g. Li & Hegelheimer, 2013; Uther, Uther, Athanasopoulos, Singh, & Akahane-Yamada, 2007), with the vast majority focusing on learners' patterns of usage and perceptions towards using the technology (e.g. Amer, 2014; Kim, Rueckert, Kim, & Seo, 2013; Stockwell, 2007, 2010; Stockwell & Liu, 2015; Tan & Teo, 2015; Wang & Smith, 2013).

Interestingly, apps aimed at teaching pronunciation are only a minority among language learning apps, revealing some kind of neglect reminiscent of the one traditionally suffered by pronunciation in FL teaching. In a review of 87 mobile apps for language learning, Kim and Kwon (2012) showed that the majority of apps concentrated on vocabulary (42.5%), followed by reading (17.4%), grammar (12.6%), listening (11.4%), speaking (9.1), and lastly writing (5.7%). Even though pronunciation is mentioned several times as 'facilitating device' in apps addressing other competences – in that apps commonly incorporate sound files to exemplify pronunciation of words or common sentences –, of the 87 apps analysed, none of them focused exclusively on pronunciation. It is important to take into consideration that the app stores from different countries do not always include the same apps. Nonetheless, it seems obvious that pronunciation as a competence is largely ignored among language learning apps, without a single app addressing pronunciation in such a comprehensive analysis.

Fouz-González (2012) presented a brief 'state of the app' discussing the potential and some limitations of a number of iPhone apps devoted to training FL pronunciation. The analysis is not half as exhaustive as the one by Kim and Kwon (2012), as the aim was simply to offer an overview of existing apps dealing with EFL pronunciation. It focused exclusively on apps from Apple's app store, and it only included those apps the researcher deemed to be potentially advantageous for pronunciation improvement.[57] As pointed out in an extended analysis also addressing Android apps (Fouz-González, 2013), a great number of existing apps do not seem to have any kind of pedagogical foundation behind – as evidenced, for example, by apps with speech synthesis fraught with inaccuracies that illustrate mispronunciations rather than a target model.

Regarding the studies addressing the potential of apps for pronunciation improvement, to the researcher's knowledge, only the studies by Uther et al. (2007) and Liakin et al. (2014) have explored the potential of a mobile application to improve the learners' pronunciation empirically. Uther and colleagues (2007) used Logan et al.'s (1991) HVPT paradigm in order to train the /r-l/ contrast to EFL learners through a Java-based application. Learners received training over a period of two weeks in which they had to use the app around 30 minutes a day. Their findings show that training with the mobile application improved the learners' perception of the contrast considerably, therefore suggesting that training with a mobile app can be as effective as traditional PC-based approaches. In a similar line, Liakin and colleagues (2014) used an ASR application to provide a group of FL learners of French with immediate feedback on the pronunciation of French /y/. Learners were required to use the app for 20 minutes per week over a period of five weeks. Their results show that the repetitions learners made of the target sentences as prompted by the feedback offered by the application led them to improve their pronunciation of /y/ as compared to other two groups, one receiving the same type of instruction in class with feedback offered by a teacher, and another one that did not receive feedback.

Social networking services: Twitter

Social networking services (SNSs) can be defined as websites or applications where users interact with each other by sending messages, posting updates, comments or pictures on their daily experiences, among others. Numerous authors have pointed out the potential of SNSs for education given learners' familiarity with these tools and the possibilities they offer for communication among students or between teachers and students, both in class and out of class. Some studies have addressed SNSs specifically created for language learning, such as Babel, Busuu, or Livemocha (Brick, 2012; Harrison, 2013; Liu et al., 2013;

---

[57] The search was conducted in the App Store under the keywords 'English pronunciation', 'English phonetics', and 'pronunciation'. Phonetics was included as a keyword because a good number of apps tend to treat phonetics and pronunciation as the same. Apps that did not include activities for practice were not included in the review, nor were those that were limited to traditional listen-and-repeat activities using ASR of questionable quality.

Stevenson & Liu, 2010), while others have decided to explore the possibilities offered by those that were not originally conceived as educational tools, such as Facebook (Blattner & Fiori, 2009, 2011; Blattner & Lomicka, 2012; Ota, 2011), or Twitter (Antenos-Conforti, 2009; Baker, 2010; Lamy & Zourou, 2013; Mork, 2009; Rinaldo, Tapp, & Laverie, 2011).

As a case in point, Facebook has been found to facilitate student-centered learning, with students taking charge of online discussions and becoming more active and confortable when sharing their opinions with other classmates (Promnitz-Hayashi, 2011). Blattner and Lomicka (2012) found that learners enjoy the 'pressure-free' and 'casual' environment Facebook offers. However, their data reveals that learners use most Facebook functions for their personal lives and limit academic use of this tool to the group discussions. This finding is consistent with other studies showing learners' reticence to share too much of their private lives in a learning site. For example, Harrison (2013) reports that many of the students who used *Livemocha* (an SNSLL) preferred not link their accounts to *Facebook* and wanted to keep their private lives and connections separated from the ones devoted to learning.

The third block of this dissertation focuses on Twitter, one of the most popular SNSs worldwide, with more than 300 million monthly users according to the official website.[58] As compared to Facebook, an SNS where users usually become 'friends' with other users in order to follow their updates (therefore granting them access to a considerable amount of personal information unless it is well customised), one of Twitter's charms is that people engange in discussions with total strangers without necessarily being subscribed to their accounts. By typing a hashtag (#) followed by any word, users can start a thread to prompt conversation with other users around the world (e.g. #Ilovetheweathertoday) and replies including the same hashtag in their posts will automatically be displayed on that thread.

One of the defining characteristics of Twitter is that 'tweets' (i.e. posts) are limited to 140 characters, forcing users to be concise in their contributions. Given the limited space allowed in posts and the type and frequency of people's posts, Twitter is often considered a form of microblogging[59] (Jansen & Zhang, 2009). People typically use Twitter to share their plans, express their views on a range of topics, or sharing different pieces of information, such as news, videos, music, etc. (Java, Song, Finin, & Tseng, 2007).

---

[58] https://about.twitter.com/company
[59] 'The practice of making short, frequent posts to a microblog' (Apple Inc., 2014).

The Web 2.0 is characterised by users being able to generate content rather than simply visiting static websites (e.g. in blogs, wikis, YouTube, or SNSs like Facebook or Twitter). This offers endless possibilities for language teaching and learning. As Lomicka and Lord (2009, p. 4) note, the Web 2.0 'is not so much about a new technological invention or platform but instead the ways we use existing technologies or platforms to do more and different things'.

For teachers, Twitter can be a platform to reflect on their teaching practices, sharing their experiences and perceptions with colleagues from different parts of the world (Lord & Lomicka, 2014), or to contact and benefit from exchanges with professionals in different fields (Dunlap & Lowenthal, 2009). For students, Twitter can be used to solve doubts or to post information about courses (Lowe & Laffey, 2011), to foster collaboration (Kassens-Noor, 2012) and communication among students (Antenos-Conforti, 2009; Borau, Ullrich, Feng, & Shen, 2009), or even help learners remember important concepts (Blessing, Blessing, & Fleck, 2012).

Focusing on language learning, some have suggested using Twitter for composition tasks, vocabulary exercises and listening activities (Mork, 2009). Even though the length of tweets is rather limited, Twitter seems an ideal tool to foster output production (Kim, Park, & Baek, 2011), enabling learners to interact with other peers as well as native speakers of the language they are learning (Liu et al., 2013). However, as with podcasts and apps, there are very few studies that address Twitter's language learning potenital empirically.[60]

A number of Twitter accounts send a 'word of the day' in order to help their 'followers'[61] learn different English words and expressions (e.g. @OxfordWords, @CambridgeWords, @MerriamWebster). Their tweets include the word in context and a link to the corresponding online dictionary, where users can find the definition of words, an audio illustration of their pronunciation, and in the case of Oxford and Cambridge, their phonemic transcription with IPA symbols (Figure 16).

[60] Language learning potential is defined as 'the extent to which the activity can be considered to be a language learning activity rather than simply an opportunity for language use' (Chapelle, 2001, p. 55).
[61] In Twitter jargon, people subscribed to their account.

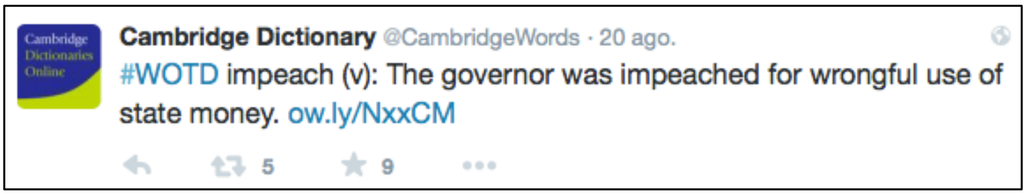


Figure 16. Sample word of the day by @CambridgeWords

Some Twitter accounts have also started tweeting about pronunciation, sharing links to the pronunciation of a range of words and sentences by different speakers with different accents (@Forvo), to YouTube videos (@PronunciationBook), or giving tips about pronunciation (@ConfidentVoice, @Forvo). Given the learners' positive attitudes towards learning vocabulary through SMS or email (Kennedy & Levy, 2008; Levy & Kennedy, 2005; Thornton & Houser, 2001, 2005), and the pronunciation gains fostered by similar approaches, such as MMS (Saran, Seferoglu, & Cagiltay, 2009), research has also addressed Twitter's potential for helping EFL learners improve their pronunciation of problematic words (Fouz-González & Mompean, 2012; Mompean & Fouz-González, in press).

The study by Fouz-González and Mompean (2012) was a pilot study that served as scaffold for a subsequent study (Mompean & Fouz-González, in press) and for study 4 in this dissertation. The approach adopted consisted in sending learners a daily tweet with explicit explanations on the pronunciation of a range of words whose pronunciation was considered to be problematic for Spanish EFL learners. The study compares the data from four students from a degree in English studies and a random selection of four students from Mompean and Fouz-González (in press). The results reveal considerable improvements for the latter group, but only moderate gains for the former. The gains correlate with the level of engagement participants showed during the instruction, which was rather low for the English studies group. A possible explanation for that difference is that the group obtaining the largest gains were only exposed to English four hours a week and were very keen on taking every chance they had to be exposed to English (see chapter VI), whereas the other group was receiving daily instruction in English in their degree and may have found the tweets to be trivial as compared to the kind of exposure they were accustomed to.

Mompean and Fouz-Gonzalez (in press) used 16 students from an Official Language School in Murcia[62] in order to test Twitter's potential to improve pronunciation as well as

[62] In Spain, Official Language Schools (*Escuelas Oficiales de Idiomas*) are a type of state-run, public language school offering language tuition and officially recognised diplomas of several foreign languages. They cover

to encourage online participation. Twitter was found to be a convenient tool for both aspects. The researchers sent a daily tweet with concise explicit explanations and audio/video illustrations on the pronunciation of 27 lexical items whose pronunciation is problematic for EFL learners. The Twitter-based instruction yielded an overall gain rate of 74.9% in the learners' post-test productions as compared to the pre-test. Similarly, student engagement in the study was quite high, with participants interacting much more actively than required by the researchers. Nonetheless, the study presents several limitations, such as a limited number of participants or the lack of a control group.

In Mompean and Fouz-González's (in press) study, six participants could not attend the post-test interviews, therefore limiting measures for production gains to 10 participants. However, despite the limited number of participants, the results are considered to offer valuable insights as regards Twitter's potential to foster pronunciation improvements, as those 10 learners obtained considerable pronunciation gains after instruction. A control group would strengthen the claim that improvements were a result of instruction. Nevertheless, given the small number of learners who volunteered to participate, and since participation was advertised as a 'free pronunciation course', no control group was used in order to avoid depriving students from instruction (see also Lord, 2008; Thomson, 2011).

CEFR levels from A1 to B2 over a period of six years. In some regions, an additional course is offered to help learners attain the C1 level. Official Language Schools (OLSs) in Spain are a very special type of education with a wide variety of student profiles. Until recently, it was one of the most common choices for Spaniards who wanted to obtain an official diploma in a foreign language irrespective of their age, often adults or university students. Thus, students enrolled in OLSs are commonly highly motivated students, as they enrol voluntarily and are willing to spend six years of their lives studying English.

Chapter II: Main objectives

2.1 General objective

Given the need for materials and techniques to support FL pronunciation instruction, and because of the limitations of current technology for FL pronunciation training discussed in chapter I, in this dissertation we investigate how affordable and accessible technologies can enhance pronunciation training in an FL context as well as the learners' perceptions towards using these tools. In particular, the studies in this dissertation deal with podcasts (chapter IV), smartphone apps (chapter V) and Twitter (chapter VI). These three tools are considered to be very accessible to learners, both in terms of their simplicity and physical accessibility.

The technologies explored here can be used both as an aid to classroom instruction or for autonomous practice. However, as noted above, the idea in this dissertation is to empower learners with self-monitoring skills so that they can notice features of the FL phonology when practising on their own. The techniques employed make pronunciation accessible for the average learner; for people who may not be especially interested in phonetics but who nevertheless need to have an adequate pronunciation in the foreign language. Finally, they represent an attempt to move away from the common laboratory-like settings where research is often conducted (one of the often-cited claims about CAPT studies, see e.g. Lord, 2010; Olson, 2014; Wang & Munro, 2004) allowing learners to use technology outside the classroom, enabling them to practise in contexts where they will ultimately benefit from these technologies for autonomous practice.

2.2 Specific objectives

The specific objectives of this dissertation relate to the potential of each of the above-mentioned technologies to improve particular aspects of English pronunciation for EFL learners.

The three studies in this dissertation are arranged in three chapters, each of them exploring the potential of one of the above-mentioned technologies to improve a particular aspect of FL pronunciation. More specifically:

In study 1 (chapter IV) we investigate:

1. The potential of podcasts to improve Spanish learners' perception of the English /s – z/ contrast and of English voiced stops /b d g/ in intervocalic position.
2. The potential of podcasts to improve Spanish learners' production of English /z/ and to help them avoid spirantising /b d g/ in intervocalic position.[63]
3. The learners' perceptions towards using podcasts for pronunciation training.

In study 2 (chapter V) we explore:

4. The potential of smartphone apps to improve Spanish learners' perception of English /æ/, /ʌ/, /ɑː/, /ə/, and the /s – z/ contrast.
5. The potential of smartphone apps to improve Spanish learners' production of English /æ/, /ʌ/, /ɑː/, /ə/, and /z/.
6. The learners' perceptions towards using smartphone apps for pronunciation training.

In study 3 (chapter VI) we address:

7. The potential of Twitter to improve Spanish learners' pronunciation of a number of problematic aspects of English often mispronounced by EFL learners due to sound-spelling correspondences or different stress patterns in cognates.
8. The learners' perceptions towards using Twitter for pronunciation training.

2.3 Research questions and hypotheses

To address the specific objectives mentioned above, the four studies in this dissertation present the following research questions (RQ) and hypotheses (H):

---

[63] The voiceless alveolar fricative /s/ is included in the perception goals in blocks 1 and 2 because learners fail to perceive the difference between this sound and its voiced counterpart /z/. However, /s/ is not a target in production given that Spanish learners of English already have this sound in their phonemic repertoire and it should not present articulatory difficulties.

Study 1 (Chapter IV)

RQ1: Can learners' perception of the English /s – z/ contrast and /b d g/ as stops in intervocalic position be improved as a result of training with podcasts?

H1: Training with podcasts will help learners improve their perception the English /s – z/ contrast.

H2: Training with podcasts will help learners improve their perception of English /b d g/ as stops stops in intervocalic position.

RQ2: Do learners' improve their pronunciation of English /z/ and /b d g/ as stops in intervocalic position after training with podcasts?

H3: Training with podcasts will help learners improve their production of /z/.

H4: Training with podcasts will help learners avoid spirantising English /b d g/ in intervocalic position.

RQ3: Does training with podcasts generalise to novel words?

H5. Instruction with podcasts will help learners improve their perception of the /s – z/ contrast in novel words.

H6. Instruction with podcasts will help learners improve their production of /z/ in novel words.

H7. Instruction with podcasts will help learners avoid spirantising /b d g/ in intervocalic position in novel words.

RQ4: What are the students' reactions towards using podcasts for pronunciation training?

Study 2 (Chapter V)

RQ1: Can instruction with an app foster improvements in the learners' perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/?

H1. Training with the app will help learners improve their perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/.

RQ2: Does perceptual training with an app foster improvements in the learners' production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/?

H2. Perceptual training with the app will help learners improve their production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/.

RQ3: Does training with the app generalise to novel contexts?

H3. Training with the app will help learners improve their perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/ in novel words.

H4. Training with the app will help learners improve their production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/ in novel words.

RQ4: Does training have the same impact on learners enrolled on an English Studies degree taking a phonetics course than on learners from other degrees not receiving training in phonetics?

H5. Training with the app will foster similar improvements in English majors taking a phonetics course than on learners not receiving training in phonetics.

RQ5: What are the students' reactions towards using apps for pronunciation training?

Study 3 (Chapter VI)

RQ1: Can Twitter help learners' improve their pronunciation of lexical items that are commonly mispronounced due to sound-spelling correspondences or different stress patterns in cognates?

H1: Sending learners a daily tweet with explicit explanations about the target aspects and audio/video illustrations will help them improve their pronunciation of the target words.

RQ2: Is Twitter-based instruction as effective when implemented with participants who volunteer to participate than with students who are offered a reward in exchange for their participation?

H2: Training with a group of students enrolled on an ESP course and rewarded for their participation will yield similar results to the ones obtained with a group of volunteer FL learners.

RQ3: What are the students' reactions towards using Twitter for pronunciation training?

2.4 Overview and rationale for the studies in this dissertation

In the same way that CALL and CAPT developments have often been technologically rather than pedagogically driven (Eskenazi, 2009; Neri et al., 2002; Pennington, 1999), some authors have criticised the higher value that is sometimes placed on the technical features of different devices rather than the pedagogical and theoretical rationale for using them (Cheung & Hew, 2009). As Jamieson, Chapelle, and Preiss (2005) point out, justification for using CALL is not as necessary today as it used to be, as computers are now used in all types of instruction. 'The question for research, then[,] is to what extent a particular type of CALL material can be argued to be appropriate for a given group of learners at a given point in time?' (Jamieson et al., 2005, p. 2). In this respect, the studies in this dissertation focus on different technologies and on different pronunciation problems, trying to exploit each tool for what it is deemed most suitable.

As noted at the end of chapter I, the three technologies explored in this dissertation are considered to be affordable and highly accessible. As far as price is concerned, the three technologies are either free (podcasts and Twitter) or relatively inexpensive (apps).[64] With regard to accessibility, learners can access them from their computers at home but also from their mobile devices, anywhere and at any time.

2.4.1 Podcasts

Podcasts are addressed in two studies in this dissertation. The first was a pilot study aimed at testing the procedure to be followed in study 1 as well as the suitability of the different tasks. Thus, although training included a selection of problematic aspects for EFL learners, the perception and production data from the pilot study have not been considered for analysis and will therefore not be reported here. Nonetheless, a detailed description of the procedure and the findings obtained from the questionnaires is offered in section 4.1 given that this led to several modifications in the method employed in study 1.

As regards the target aspects addressed in study 1, two issues that are known to be problematic for Spanish learners of English were selected for training, namely the pronunciation of English voiced stops /b d g/ in intervocalic position and the English /s – z/ contrast.

---

[64] The price for an average language learning app ranges from being completely free up to approximately 5€. To offer an idea on app prices, of the pronunciation apps explored in Fouz-González (2012), the most expensive were the *Sounds* app (Macmillan Publishers Ltd., 2011) – costing 5.99€ – and the *English File Pronunciation* app (Oxford University Press, 2012) – costing 5.49€. However, there are more affordable apps, such as the *iBaldi* talking head (Psyentific Mind, Inc., 2012) – which costs 1.99€, or the *Sounds of Speech* app (UIRF, 2014) – which costs 2.99€. It should be noted that prices are subject to change. The *Sounds* and *English File Pronunciation* apps cost 4.99€ in 2012 (Fouz-González, 2012), but now cost 5.99€ and 5.49€ respectively. These prices have been taken from the Apple's app store in September 2015.

a) Target aspects addressed

The English /s – z/ contrast

Spanish learners of English often fail to perceive and mark the distinction between English /s/ and /z/(see e.g. Hazan & Simpson, 2000; Magen, 1998), probably due to equivalence classification (Flege, 1987) as /z/ does not exist in the Spanish consonantal inventory. Although a voiced alveolar fricative did exist in Medieval Spanish (Hualde, 2014), and even though /s/ may sometimes be realised phonetically as /z/ due to assimilation processes when followed by a voiced sound (Hualde, 2014; Martínez-Celdrán et al., 2003), Castilian Spanish only has one alveolar fricative in its phonemic repertoire (Hualde, 2014; Quillis, 2012), the voiceless /s/. In this respect, four of the English sibilants /s z ʃ ʒ/ that do not have phonemic value in Spanish tend to be merged into the Spanish phonetic category for /s/(Monroy, 2001). Some of the examples offered by Monroy are *English* [ˈɪŋglɪʃ] pronounced as *[ˈiŋglis];[65] *occasion* [əˈkeɪʃn̩] as *[oˈkeison]; *noises* [ˈnɔɪzɪz] as *[ˈnoisis]; *because* [bɪˈkɒz] as *[biˈkos]; or *usually* [ˈjuːʒʊəli] as *[ˈjuːsuali].

As with issue the pronunciation of English /b d g/ in intervocalic position, despite the fact that mispronunciation errors with /s/ and /z/ are not likely to cause intelligibility problems except for minimal pairs (such as *sip – zip*; *precedent – president*; *bus – buzz*), it was considered to be an interesting issue to address with Spanish-speaking EFL learners given that it tends to be fossilised in the interlanguage of very advanced learners (Monroy, 2001).

The pronunciation of English /b d g/as stops in intervocalic position

The pronunciation of English voiced stops /b d g/ is known to be problematic for Spanish learners because they tend to pronounce these three phones as approximants in intervocalic position as a result of first language transfer (Schairer, 1992; Zampini, 1994, 1996, 1997). English and Spanish have the same three voiced stops in their phonemic repertoires, namely /b d g/ (see e.g. Cruttenden, 2008; Roach, 2000 for English, or Hualde, 2014; Quillis, 2012 for Spanish). However, the pronunciation of these sounds poses problems for both types of learners, English-speaking learners of Spanish and Spanish-speaking learners of English as a foreign language.

[65] The other vowels are also represented with the IPA symbols for Spanish because they were also mispronounced by Monroy's informants.

While English /b d g/ are normally realised as stops irrespective of their phonetic environment,[66] their Spanish counterparts are spirantised in intervocalic positions, rendering three approximant allophones [β ð ɣ] respectively (Martínez-Celdrán, Fernández-Planas, & Carrera-Sabaté, 2003; Hualde, 2014). Although in some dialects of Spanish /b d g/ are realised as stops in other positions (see Zampini, 1996), in Peninsular Spanish, the variety spoken by participants in this study, they are only pronounced as stops in word initial position after a pause, after a nasal consonant (as in *bombo, cuando,* or *mango*), or after /l/ in the case of /d/ (as in *molde* or *caldo* – see Hualde, 2014). Hence, English learners' mispronunciation of these phones when speaking Spanish consists in their realisation as stops in every context (Lord, 2010; Olson, 2014; Zampini 1994, 1997), whereas the problem for Spanish learners is that they tend to transfer their L1 spirantisation rule to English (Zampini, 1996).

The difference in pronunciation between English and Spanish voiced stop consonants /b d g/ has long attracted researchers' attention. However, while this is a problem that affects both English-speaking learners of Spanish and Spanish-speaking learners of English, investigations have often centred on the acquisition of these sounds by English learners of Spanish (e.g. González-Bueno, 1997; Lord, 2010; Olson, 2014), with only a few focusing on Spanish EFL learners (e.g. Rochdi & Mora, 2012).

Despite the fact that mispronunciation of /b d g/ as approximants in intervocalic position is not among the factors that most affect intelligibility (Schairer, 1992), it can be considered to be one of the factors that exhibits noticeable traits of Spanish-accented speech. As noted in chapter I, while many foreign language learners may be content with sounding 'comfortably intelligible', many others wish to attain native-like proficiency in English. In this respect, participants in the studies in this block were enrolled in a degree in English studies. They are considered to be the type of students who may, in theory, be very motivated to learn the language – given that they have decided to spend four years of their lives specialising in one or more foreign languages, often with the intention to teach those languages to FL learners.[67] Hence, it is assumed that many of these learners may wish to

[66] Voiced plosives may occasionally be realised as fricatives in 'very rapid, familiar speech, where speed rather than articulatory precision is the aim' (Cruttenden, 2008, p. 168). Examples of this for /b/ and /g/ are the pronunciation of *rubber* as [ˈrʌβə] or *dagger* as [ˈdæɣə] (*ibid.*). Similarly, in American English /d/ may be realised as a flap allophone [ɾ] when preceding unstressed syllables, such as in *caddy* [ˈkæɾi] or *ladder* [ˈlæɾə] (Celce-Murcia et al., 2010) – a tendency that is also observable in Southern Standard British English (SSBE), although it is much more common for /t/ than for /d/ (Shockey, 2003).

[67] The questionnaires for studies 1 and 2 do not include items addressing the learners' intended careers. Nonetheless, this degree in English Studies is the degree that people have traditionally studied in order to

sound like native-speakers, and in fact, the responses to the questionnaires in studies 1 and 2 confirm this prediction.

The data offered by Monroy (2001) attest to the difficulty of Spaniards in pronouncing /b d g/ adequately in intervocalic position, as it is an aspect often fossilised in the interlanguage of very advanced learners of English. This suggests that, regardless of the impact of mispronouncing /b d g/ as approximants on intelligibility, it is something particularly difficult to learn for Spanish speakers. Informants in Monroy's study were third-year students enrolled in the former English Philology degree. Students who hold this degree are supposed to have a C2 level according to the Common European Framework of Reference for Languages (CEFR),[68] and after they complete an additional pedagogical component, they are qualified to teach English in secondary education and Official Language Schools in Spain. In this regard, participants in studies 1 and 2 had a similar profile, as they were enrolled in the second year of the new English Studies degree (a four-year degree that substitutes the former English Philology).

b) Rationale for using podcasts

Podcasts provide an auspicious solution to the scarcity of available input in FL contexts, one of the greatest limitations FL learners face in comparison to children learning their mother tongue or learners of a second language. Podcasts not only bring a virtually unlimited amount of authentic input to the learners' fingertips, but they also offer a wide variety of topics and programmes of various lengths, featuring a rich diversity of accents – something very convenient for pronunciation practice. This makes it possible for teachers and learners to use materials that satisfy different interests, while matching their time needs and possibilities.

Although finding authentic materials was difficult and sometimes expensive until recently (e.g. cassettes and CD-ROMs), today's learners can download any podcast of their

---

become teachers of English as a FL in Spain (formerly known as *English Philology,* a five-year degree covering a wide range of subjects on English linguistics, translation, literature, history and culture). It is assumed that people who want to become language teachers in a foreign country should aspire to the highest proficiency possible. In fact, the students' responses to the questionnaires in studies 1 and 2 reveal a mean score of 4.4 in a five-point Lickert scale question (SD = 0.8; n = 101) asking them about their willingness to acquire native-like proficiency.

[68] Even though not all students who complete this degree have a genuine C2 level, there are several official documents that equate the degree in English Philology or a degree in Translation to the Cambridge Proficiency in English by Cambridge University (i.e. the C2 level according to the CEFR). See for example: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-5881.

choice anywhere thanks to smartphones. This offers countless possibilities for language learning in general and for pronunciation practice in particular.

As discussed in chapter I, despite the vital importance of input in foreign language acquisition (Krashen, 1982, 1985), output has also been identified as essential given the different functions it fulfils, such as helping learners to notice gaps in their knowledge/skills, testing their hypotheses or gaining metalinguistic skills to reflect on language (Swain, 1985, 2005). With regard to pronunciation, input is paramount inasmuch as a new phonological system is only acquired after being exposed to numerous instantiations of the target sounds and patterns (Flege, 1991, 1995). Nevertheless, output is also central to pronunciation acquisition, as learners need to automatise the articulatory movements of the FL, often different from those of their mother tongue (Neri et al., 2002). Podcasts bring these two elements to the learners' fingertips, giving them access to an incredible amount of authentic natural input of virtually any accent and topic, and allowing them to record and share their own output in podcast format with other people – ideal in order to benefit from the three functions of output and work on their self-monitoring skills.

The idea in this study is to help learners notice relevant features in the input they are exposed to by means of explicit instruction. Hopefully, this should eventually lead to the creation (or consolidation) of adequate phonetic categories for the target FL sounds, which are hypothesised to guide their subsequent productions (see discussion in section 1.1.5 above). Furthermore, podcasts are also used in order to foster opportunities for consolidating articulatory movements and self-monitoring, as learners are also asked to record their own podcasts and evaluate their pronunciation and their peers' (see the hypotheses entertained below the research questions in section 4.2.1).

A case was made for the importance of FL phonological awareness and self-monitoring in pronunciation in section 1.1.5. Podcasts seem an ideal tool in order to foster both. The way learners' awareness of the FL phonology is raised is in line with several SLA concepts, namely focus on form (Long, 1991), explicit instruction (Saito, 2013), input enhancement (Sharwood-Smith, 1993), noticing (Schmidt, 1990, 2010), and the three functions of output (Swain, 1985, 2005). Although the details of the approach will be explained further in chapter IV, a brief overview is provided in this rationale in order to establish the link with the aforementioned theoretical concepts.

In the approach adopted, learners are first given a brief explicit explanation of the problematic target aspects with the intention of fostering their noticing of those problematic features when they hear them in the input. Next, input is considered to be ‘enhanced’ by the metalinguistic information provided to learners (Ioup, 1995), but also through the focused listening activities in the second stage of the process (activities 1 and 2). Even though input is not visually or acoustically modified in any way, explicitly directing learners’ attention to specific words and phonemes could act as some kind of ‘enhanced listening’ – as learners have to pay attention to features that may otherwise go unnoticed. It is at this second stage where participants are expected to notice features in the input that they should subsequently apply in their productions. At the third stage, learners are encouraged to record their own podcasts, which activates the three functions of output outlined by Swain (1985, 2005). By attempting to articulate the target sounds, learners may notice ‘gaps in their knowledge’ – in the case of pronunciation, their inability to articulate them as expected, realising that they cannot pronounce the target adequately (the noticing/triggering function). Then, their attempts at pronouncing the text in the podcast are a way of testing their hypotheses about the articulation of those sounds, making use of their mental categories for such sounds (the hypothesis-testing function). Finally, when learners are required to evaluate their peers and offer them feedback, ideally, explicitly indicating them how to improve, they are required to reflect on and talk about language (the metalinguistic function). The method is illustrated visually in Figure 17.

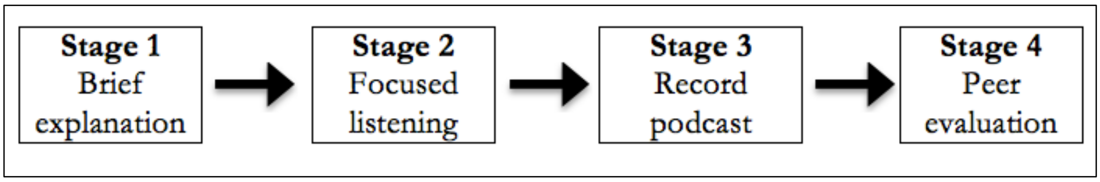


Figure 17. Diagram of the different stages in the podcast-based instruction

This approach is in line with recommendations for self-monitoring in the literature. As Avery and Ehrlich (1992, p. 218) suggest, techniques for self-monitoring may include:

> listening to a taped sample of speech and identifying appropriate and inappropriate forms […] determining whether pronunciation has been accurate […] By contrasting good with bad habits, the instructor allows students to experience the difficulties which listeners may have in conversation with them. Peer feedback can also convince students that certain areas of their speech need practice.

In fact, the results obtained by Baker and Trofimovich (2006) suggest that self-perception abilities play a crucial role in the development of adequate L2 productions. In their study, despite the fact that some learners' production accuracy was better than their pereptual accuracy, self-perception was always better than production. They conclude that perceiving their speech accurately was essential for them to articulate the target sounds accurately. In their words, 'without self-perception, which may occur at an auditory (Flege, 1995) or articulatory (Best, 1995) processing level, L2 learners may not be able to link auditory perceptual targets with their own articulatory efforts' (Baker & Trofimovich, 2006, p. 247)

The method proposed here is similar to other recommendations that promote the learners' noticing of features in the input and analysis of their own output, such as the oral readings used by Tanner and Landon (2009) or Walker (2005), the critical listening recommended by Fraser (2001) or Couper (2011), or the 'Noticing-Reformulation technique' put forward by Smith and Beckmann (2005, 2010).

In the approach adopted by Smith and Beckmann's (2005, 2010), learners are first asked to record a text and analyse their own speech paying attention to specific features. Then, they are given a model for comparison in order to foster noticing of the model's pronunciation of the target features – with the intention of helping them 'notice the gap' between their pronunciations and the model, which will hopefully lead to an improved version of the first try. Finally, learners compare their two recordings in order to reflect on possible changes.[69]

The method adopted here is more similar to the one used by Tanner and Landon (2009) or Walker (2005) in that the model is presented first, encouraging learners to notice features in the input so that they can implement them in their output. Given the perceptual challenges FL learners face with segmental features, the activities in block 1 placed greater emphasis on the learners' identification of the target sounds in the input. This was done with the intention of fostering learners' implementation of features they perceive in the input to their developing target categories for the FL sounds (see Saito, 2013). By consciously paying attention to the way sounds are pronounced, learners are hypothesised to create or consolidate their FL phonetic categories (Guion & Pederson, 2007), as they will be exposed to different instantiations of the same sound pronounced by different speakers and in different phonetic contexts (see Flege, 1991).

[69] For a more detailed description of the process see Smith and Beckmann (2010, pp. 37-38).

One of the negative aspects of CAPT courseware using ASR noted by O'Brien (2006) is learner autonomy, which tends to be rather limited in some ASR systems since learners can only repeat those sentences that are pre-stored in the system. Once learners are ready to self-monitor their pronunciation, the approach proposed with podcasts will enable them to practise with any podcast they want, anywhere, and with the accent of their choosing. The idea is to empower learners to start paying attention to features of the input they are exposed to in an attempt to help them subsequently implement those features in their production. This is considered particularly helpful for students, as it resembles the way students are naturally exposed to FL input. The benefits could be extrapolated to their daily practice once they became capable of monitoring their productions as well as those of others.

2.4.2 Smartphone apps

As noted in section 1.3.3, apps are self-contained programmes downloaded into a mobile device (e.g. tablets, smart watches or smart TVs) and that can be accessed at the users' convenience, both in terms of place and time. Study 2 in this dissertation focused on smartphone apps (rather than tablets or any other device) because smartphone ownership was considered to be fairly widespread among students – this was indeed confirmed by the questionnaires in studies 1 and 3 (see below).

Numerous studies describe the benefits obtained from perceptual training in perception (e.g. Carlet & Cebrián, 2014; Rato, 2014) and production (e.g. Bradlow et al., 1997; Lambacher et al., 2005; Thomson, 2011; Wong, 2015). Hence, study 2 in this dissertation explores the potential of perceptual training provided by a smartphone app on the learners' perception and production of four vowel and two consonant English sounds often problematic for Spanish learners of English.

a) Target aspects addressed

The target items addressed in this study are four English vowels that Spanish learners of English commonly merge into their /a/ category, namely /æ/, /ɑː/, /ʌ/ and /ə/ (see Fox, Flege, & Munro, 1995; Monroy, 2001), and the contrast between /s/ and /z/, given that Spanish does not have /z/ and is often heard and pronounced as/s/. Since the /s – z/ contrast was already explored in section 2.4.1 above, only the target vowels will be addressed here.

The British English sound system (the target model in studies 1 and 2 in this dissertation) has 12 monopthongs, namely /iː ɪ ɛ ə ɜː æ ʌ ɑː ɒ ɔː ʊ uː/, whereas the Spanish system only has five, /a e i o u/. While Spanish vowel sounds like /i/, /o/, or /u/ have only two potentially problematic English counterparts – in that they are phonetically similar and subject to equivalence classification – (/iː/ and /ɪ/ for /i/; /ɒ/ and /ɔː/ for /o/; and /ʊ/ and /uː/ for /u/), there are four English monopthongs that can be merged into the Spanish category for /a/, as noted above: /æ/, /ɑː/, /ʌ/ and /ə/. This is not only problematic in terms of perception and articulation of the target sounds, but also given their various possible orthographic representations – which make the task of knowing when to pronounce each sound extremely daunting for learners.

The examples of vowel substitutions offered by Monroy (2001) pertain to vowel quality, such as English /æ/ for Spanish /a/, as in *family* *[ˈfa̲mili]; English /ə/ and /ʌ/ for Spanish /a/, as in *another* *[a̲ˈna̲ða̲r]; and vowel duration, such as English /ɑː/ for Spanish /a/, as in *castle* *[ˈkasel] or *marvellous* *[ˈmarβelus]. Nevertheless, although Monroy's data were obtained form a spontaneous task where reading could not possibly influence the learners' pronunciation, the influence of orthography is clearly noticeable in the learners' mental representations of the target sounds, as evidenced in pronunciations such as *[ˈfami̲li] for *family* (/ˈfæmə̲li/), [po̲ˈlisma̲n] for *po̲licema̲n* (/pəˈliːsmən/), or *[blo̲d] for *blo̲o̲d* (/ˈblʌd/).[70]

b) Rationale for using smartphone apps

Phones have been considered as one of the 'basic necessities of life' in our culture as a result of their ubiquity (Kenning, 2007, p. 171). The increase in mobile phone ownership is manifested in several studies reporting all participants owning mobile phones (e.g. Stockwell, 2013; Thornton & Houser, 2005), although the numbers for smartphones have not been as promising depending on the country and year when the studies where conducted (see e.g. Stockwell, 2012, 2013; Stockwell & Liu, 2015). In Spain, the context where all studies in this dissertation were conducted, 81% of mobile phones are smartphones (Fundación Telefónica, 2015).[71] In fact, as revealed by the participants' responses to the questionnaires in this dissertation, of the 241 respondents in the initial questionnaires to studies 1, 2 and 3, 92.5 per cent owned a smartphone.

---

[70] Spanish vowel sounds are always represented in orthography by the phonetic symbol coinciding with that vowel (e.g. *ca̲sa̲* /a/, *pe̲lí̲cu̲la̲*, /e/, /i/, /u/ and /a/, or *co̲che̲* /o/ and /e/).
[71] http://www.abc.es/gestordocumental/uploads/internacional/siE2014.pdf

Some authors have suggested that given this increase in the use of mobile devices, learners will welcome course materials delivered to their devices (Ally, 2007). However, several studies suggest that when learners are given the choice between desktop computers and mobile devices, they do not always opt for the latter (Stockwell, 2010; Stockwell & Liu, 2015). Kukulska-Hulme (2005) highlights the fact that despite the fact that these devices are perfectly suitable for the functions for which they were devised, they may sometimes be inadequate for educational contexts.

A key issue in this respect is usability, 'in its most basic definition: whether something can be used for its intended purpose' (Kukulska-Hulme, 2005, p. 45). One of the common limitations cited when evaluating mobile phones for language learning purposes is keyboard and screen size (Chinnery, 2006; Daniels, 2010; García Cabrero, 2002; Godwin-Jones, 2011; Levy & Stockwell, 2006). In terms of usability, while mobile phones can be perfectly effective for sending short messages or making phone calls, they may not be as appropriate for reading long texts. Nonetheless, this may no longer be a problem with today's smartphones, as keyboards and screens now share a common space that is becoming increasingly bigger.[72]

Focusing on the use of smartphone apps for language learning and, in particular, for pronunciation training, as noted in section 1.3.3, while numerous studies have addressed the learners' usage and perceptions towards using this technology, very few studies have investigated its potential empirically. As pointed out by Swan (2003, p. 14),

> learning effectiveness must be the first measure by which online education is judged. If we can't learn as well online as we can in traditional classrooms, then online education itself is suspect, and other clearly critical issues, such as access, student and faculty satisfaction, and (dare we say it) cost effectiveness are largely irrelevant.

That is, even though learners' preferences and learning styles are vital in the teaching-learning process, there is no point in asking them what they think about a particular technology if it does not have any real potential for language acquisition.

Mobile learning shares most of the advantages mentioned for CALL in section 1.2, such as the possibility to receive individualised instruction, immediate feedback, and practise at one's pace. The added value with mobile learning is portability. Although apps

[72] The Nokia 3310, one of the most common models in the year 2000 had a screen size of 1.8 inches; the iPhone 3GS (released in 2009) 3.5 inches; the iPhones 5C and 5S (released in 2013) 4 inches; and the new iPhone 6 (released in 2014) was made available in two different sizes, 4.7 and 5.5 inches.

can also be used in class to enhance presentation of certain aspects (e.g. articulatory movements with an animated head) or to provide individualised practice, one of their greatest assets is that learners can take them outside their classroom and practise whenever and wherever they want. Given the lack of materials to practise pronunciation and the neglect often suffered in FL programmes, smartphone apps offer promising prospects for learners who may need extra practice in an environment of their choosing.

Notwithstanding the above, and despite apps' potential for language learning and their low prices, some authors report that learners are not willing to spend money on language learning apps, as they do not trust the available commercial apps, sometimes claiming that content is not helpful or significant (Stockwell, 2012; Stockwell & Stockwell, 2012). The original intention in this dissertation was to develop an app and test its effectiveness in the block devoted to apps. However, because of the impossibility to release the app in time, one of the current existing apps developed by Oxford University Press was selected to test its effectiveness empirically, namely the *English File Pronunciation* app (OUP, 2012) – henceforth *EFP*. The app was selected on the grounds of the contents that it taught (based on the assumptions presented in section 1.1.5) as well as its availability for the different smartphone models.

Of the pronunciation apps reviewed in Fouz-González (2012), three apps were considered potentially useful to train different problematic aspects of English pronunciation because of the amount of practice offered. These were *Sounds* (Macmillan Publishers Ltd., 2011), the *EFP* app (OUP, 2012), and the two versions of *Clear Speech* (Cambridge University Press, 2011).

The *Sounds* app offers interesting possibilities to help learners become familiarised with the phonological system of English, including transcription practice. However, since it is only available in a limited range of smartphones, it was not considered convenient given that students were going to use their own phones and this would raise compatibility issues.[73] *Clear Speech* was considered promising in order to help learners work on lexical stress, rhythm and vowel reduction, or the pronunciation of final consonants in a very friendly and game-like manner. This app would have been a good candidate given the

---

[73] This app was available for all iPhone models, but only for a comparative small range of Android phones. The responses to questionnaires from studies 1, 2 and 3 revealed that of the 169 respondents who replied to this question, only 14.2% had iPhones. The majority used Android systems (84.6%) and only a minority used a differen system (1.2%). Thus, this app was discarded.

importance of lexical stress for intelligibility (Field, 2005) and comprehensibility (Hahn, 2004). Nonetheless, it had been discontinued at the time of the study.

There were other three apps not included in the above-mentioned state of the app that were also considered to offer interesting opportunities for perceptual training, such as *English Ear Game* (Kazuya Kamioka, 2011a) and *English Ear Game 2* (Kazuya Kamioka, 2011b), or *King of Vowels* (Information Technology Ltd., 2014). Nevertheless, these were not viable options either, as *King of Vowels* had also been discontinued at the time of the study and *English Ear Game* and *English Ear Game 2* addressed the /r/-/l/ and the /b/-/v/ contrasts respectively. Similarly, despite the fact that apps such as the *Sounds of Speech* app (University of Iowa Research Foundation, 2014) or *Pronunciation Power* (English Computerized Learning Inc., 2010) are potentially useful in order to illustrate certain aspects of English pronunciation, they were not contemplated as candidates for study 2 because they do not include any kind of practice.

The pedagogical rationale behind the selection of the *EFP* app (OUP, 2012) is that by helping learners become familiarised with the phonemic inventory of English and its symbols, they would be better equipped to 'label' (and consequently classify) future instantiations of the sounds they hear. The idea underlying the study is that perceptual training can lead to improvements in production, and more specifically, that teaching learners the phonemic symbols of sounds that are problematic for them will be very advantageous in order to help them create, strengthen, or correct their mental representations for those sounds. As explained in section 1.1.5, if learners do not have any referent for /æ/, /ʌ/, or /ɑː/, they will classify instantiations of these three sounds under the most similar vowel category they have in their native language, namely /a/. However, providing them with three different lables for those three sounds and helping them become familiarised with different examples of those sounds (in different positions, as pronounced by different voices and under different spellings) should help them categorise future instantiations of those sounds better. Ideally, this would strengthen their mental categories for those sounds (hypothesised to lead their subsequent productions) and consequently help learners monitor their performance.

In light of the above-mentioned linguistic assumptions, the *EFP* app was considered adequate for the purposes of the study. The app helps learners become familiarised with the sound system of English by using phonemic symbols and providing users with different instantiations of English sounds in different positions, under different spellings. Training consists of two activitities that offer immediate feedback on the learners' choices.

The activities foster the learners' familiarisation of English sounds with their phonemic symbols, provide aural identification practice of English phonemes, and test the learners' knowledge of the phonological make-up of common words featuring those sounds (see section 5.2.2 for details).

2.4.3 Social networking services: Twitter

Finally, study 3 in this dissertation (chapter VI) explores Twitter's potential to help learners improve their pronunciation of a number of problematic words commonly mispronounced by Spanish learners of English. The aspects addressed are considered to be problematic not because of perceptual or articulatory difficulties, but because of lack of exposure to the target items, difficult sound-spelling correspondences, lack of conscious attention to them (i.e. noticing), explicit instruction, or a combination of these.

The study in this block intends to replicate findings obtained in similar studies (Mompean & Fouz-González, in press), although with a different type of student profile. Participants in Mompean and Fouz-González (in press) were a group of highly motivated learners who offered to participate in the study as something completely voluntary, as something additional to their EFL lessons in an Official Language School (see section 1.3.3). The study in chapter VI addresses Twitter's potential by using a different type of learner profile, namely students enrolled in an ESP course as part a their degree in Medicine. Even though students in this study also volunteered to participate, they did not do it out of intrinsic motivation. They were awarded 0.3 points of the final mark in exchange for their participation. Additionally, study 3 intended to overcome some of the limitations in Mompean and Fouz-González's study, such as the lack of a control group or a delayed post-test.

a) Target aspects addressed

The focus was on two broad types of problems, more specifically: 1) incorrect sound choices as a result of orthography and 2) incorrect stress placement due to English and Spanish cognates with different stress patterns. With regard to the inappropriate selection of sounds, errors commonly result from: a) unusual sound-spelling correspondences in English (e.g. /v/ instead of /f/ for <ph> in *Stephen*); b) associations of English graphemes/digraphs with the Spanish phonemic value, rather than the English one (e.g. /tʃ/ instead of /ʃ/ in *charlatan*, given that <ch> in Spanish is always /tʃ/, as in *bicho* /ˈbitʃo/); and c) letters in orthography that do not materialise in pronunciation (e.g. <s> in *aisle*, <l> in *talk*, or <w> in *Greenwich*), as silent letters are unusual in Spanish. Concerning inadequate stress placement, errors are considered to derive from unexpected lexical stress in English and Spanish cognates with different stress patterns (e.g. *catholic-católico; politics-política*). For a more comprehensive account of this type of errors including examples see Mairs (1989) or Monroy (2001).

Sounds that are mispronounced in the target words have near equivalents in the learners' L1 (despite subtle phonetic differences between English and Spanish) and, therefore, learners should not need great perceptual or articulatory adjusting. That is, if Spaniards can pronounce /k/ correctly in a word like *arca* /ˈarka/ in Spanish, they should not have problems pronouncing it in a very similar phonetic environment in English (e.g. *archives* /ˈɑːkaɪvz/), or avoid pronouncing a sound for silent letters. Likewise, modifying stress within a word should not pose problems for Spaniards, as Spanish is a variable-stress language and words can be accented in any position in their L1 (see e.g. Ashby & Maidment, 2005).

b) Rationale for using Twitter

The motivation for using Twitter for pronunciation instruction comes from its conciseness (140 characters per tweet), the possibility to share audio/video files freely, and its high accessibility. Twitter is potentially very attractive given that it allows the sharing of succinct, explicit pronunciation tips to draw the learners' attention to aspects that are potentially problematic in the FL without taking much time from the learners' busy schedules. It allows teachers to illustrate the target pronunciation points by sharing audios/videos that exemplify the pronunciation problems in authentic contexts without implying any cost for either teachers or learners.

Similar technologies are limited to 160 characters per message, such as SMS or MMS. The former are restricted to text, while the latter also allow sending pictures and audio files. However, instruction using this technology would be expensive. In Spain, SMS used to cost around 15 cents and MMS around 60 cents per message. Twitter is potentially very promising in that it is completely free and it allows audio/video sharing, something extremely convenient for pronunciation instruction.

Furhtermore, Twitter can be accessed on a variety of devices (computers, smartphones, tablets, etc.), enabling learners to read the teacher's tips anywhere. Hence, this allows learners to make the most of their time by filling in time gaps between other tasks (e.g. while waiting for or travelling on the bus) or combining them with tasks that do not demand high concentration with 'bite-size' learning – one of the clearest advantages of mobile learning (Kenning, 2007).

As noted in section 1.3.3, the study was also inspired by the existence of different online dictionaries and Twitter accouts that tweet a 'word of the day' focusing on vocabulary or pronunciation. Since research suggests that vocabulary can be taught effectively via SMS (Kennedy & Levy, 2008; Levy & Kennedy, 2005; Thornton & Houser, 2001, 2005) or MMS (Saran, Seferoglu, & Cagiltay, 2009), and given that the target stimuli in this block consisted of individual lexical items, Twitter should be equally or even more effective than SMS, as it does not imply any cost for either teachers or learners and it allows audio/video sharing. Moreover, since most students already use Twitter (as indicated in their responses to the questionnaires in study 3), they can benefit from this type of bite-size learning by reading the researcher's tweets (or their teacher's – if implemented in a regular course) during their e-routine (i.e. as they do their routine check of the other accounts they follow).

In addition, as Levy and Kennedy (2005) note, mobile devices and vocabulary learning are very well suited, as research on memory suggest that spaced repetition in vocabulary learning is benefitial for robust learning. In their words:

> The deeper reasons for the likelihood of a match between mobile technology and vocabulary learning lie in well proven findings from memory research in general (Baddeley 1990) and second language vocabulary learning research in particular (Bloom and Shuell 1981). Research shows unambiguously that spaced repetition in vocabulary learning results in more robust learning than massed repetition (Nation 2001: 76). In other words, repetitions undertaken across a period, usually at ever-increasing intervals, are a more effective way to learn and retain new words than sustained repetition during a single, continuous period. SMS messages sent at intervals via a mobile phone have the potential to meet this requirement rather well. (Levy & Kennedy, 2005, p. 77)

Although tweets in this study were not repeated over time, it was hypothesised that devoting one tweet to one target item a day (i.e. devoting one 'lesson' to just one item at a time) would increase the likelihood of learners remembering the tip better. If words are retained better when learnt in isolation as opposed to mass learning, it could be the case that the learners' retention of a problematic word's pronunciation was facilitated by directing their attention to just that word – as opposed to simply listening to the word in the input amidst plenty of other words occurring in a stream of speech.

Chapter III: Methodology

The three studies in this dissertation follow a mixed-methods design, with pre- and post-tests aimed at evaluating the learners' pronunciation of the target aspects before and after receiving instruction. Additionally, questionnaires were administered at the beginning and at the end of each study in order to gather demographic data from participants, familiarity with the tools used, as well as to canvass their opinions towards the approach adopted and the adequacy of each tool for pronunciation training.

The following sections describe the general characteristics of the instruments, stimuli, and evaluation for the perception and production tasks for the three studies in the dissertation. However, the specific stimuli used and the procedure adopted in each study will be presented in chapters IV, V and VI.

3.1 Instruments

Studies 1 and 2 used pre- and post-tests aimed at assessing the learners' perception and production of the target aspects. Study 3 only evaluated the learners' pronunciation of the target aspects, as exposing them to the targets perceptually could have training effects and would not allow measuring whether potential improvements were a result of instruction. Since studies 1 and 2 follow a similar structure, they are presented together. Study 3 is presented separately.

3.1.1 Stimuli

Testing stimuli for the perception tasks were obtained from several English dictionaries as well as specialised English pronunciation dictionaries, with the exception of 13 examples of plurals in study 3 (/s/ vs. /z/) that were not included in the audio illustrations offered by pronunciation dictionaries.

Seven dictionaries were used for the selection of test stimuli. The two main sources were the Cambridge Pronouncing Dictionary (Jones et al., 2011) and the Longman Pronunciation Dictionary (Wells, 2008). A number of English learner dictionaries were also used. These were the CD-ROM version of the Oxford's Advanced Learner's Dictionary

(Hornby, 2010) and a range of online dictionaries, including MacMillan, Oxford, Collins, and Word Reference.[74] Including recordings from different speakers has two important advantages: on the one hand, it facilitates the creation of a task with a considerable amount of unique tokens, and on the other hand, it yields speaker-independent measure of participants' perception (Baker & Trofimovich, 2006). Audio files were downloaded using a trial version of Total Recorder (High Criteria Inc., 2013), a programme that allows capturing audio files as they are reproduced by the computer.

Even though the audio files were obtained from commercial dictionaries and were therefore professionally edited, they were normalised for peak amplitude (-5db) using Audacity. Participants listened to the stimuli over RoyCan RC-150D headphones in an individual computer in the computer room at the University of Murcia premises. They could adjust the volume of the stimuli presented according to their needs. In the discrimination tasks, inter-stimuli intervals were 1s long. The odd item out occurred with equal frequency in all positions, although in some cases, there was a higher number of 'catch triads' depending on the total number of stimuli (for a detailed description of the procedure for each study, see sections 4.2.2 and 5.2.2).

With regard to the training stimuli, they were always obtained from already-existing materials. In study 1, training stimuli were obtained from the '6 minute English' podcasts series; in study 2, they consisted of the audios included in the app; and in study 3, they were a number of videos and audios obtained from different internet sources.

It should be noted that in a great deal of phonological acquisition studies, both training and testing materials are created by the researchers, reserving either certain speakers or phonetic contexts for pre- and post-tests in order to test the participants' transfer of learning to new talkers or contexts (e.g. Logan et al., 1991; Lively, Logan, & Pisoni, 1993; Thomson, 2011). In studies 1 and 2 in this dissertation, generalisation was tested by comparing the participants' performance in words that appeared during training to their performance with words that did not, as the input received (in terms of speakers' voices) was always different in training and in testing.

[74] MacMillan (http://www.macmillandictionary.com), Collins (http://www.collinsdictionary.com), Oxford (http://www.oxforddictionaries.com) and Word Reference (http://www.wordreference.com).

3.1.2 Perception tasks

Perception of the English /s – z/ contrast and of /æ ʌ ɑː ə/ was measured with two common perceptual tasks aimed at assessing the learners' perceptual categories for the FL sounds, namely a discrimination task and an identification task.

The reason for the inclusion of two different tasks was that each one provides information on different perceptual abilities. One of the defining characteristics of discrimination tasks is that they allow for inter-stimuli comparisons (Beddor & Gottfried, 1995). Listeners can rely on the physical similarities/differences between sounds, whereas identification tasks require learners to resort to the mental representation they have for each particular sound. A discrimination task requires listeners to simply decide whether they perceive the stimuli presented as being the same or different. For example, if the English voiceless and voiced alveolar fricatives /s/ and /z/ tend to be perceived by Spaniards as belonging to their native category for /s/, good performance in a discrimination task would indicate that listeners are able to perceive differences between these two sounds, even if they do not know which is which. Put simply, were they to mark stimuli in a minimal pair such as *sip-zip* as different, at least they would know that what they are hearing 'is not the same'. This is especially advantageous for learners who may not be familiar with phonetic symbols given that the task does not require labels. An identification task, on the other hand, requires listeners to label stimuli individually, without having any other stimuli for comparison.

The discrimination tasks employed in this dissertation are known as oddity discrimination tasks (see Beddor & Gottfried, 1995; Logan & Pruitt, 1995). Stimuli are presented in blocks of three minimally paired words that either have the same phonological composition (i.e. 'catch triads' – *cat-cat-cat*), or one of them has a different phoneme and learners have to identify which (i.e. 'change triads' – *cat-cat-cut*). The 'odd' stimulus can appear in any position, requiring learners to mark where by clicking '1', '2', '3', or choose 'the same' if there is no odd stimulus and the three items are phonemically identical. In order to avoid response biases and random responses, an 'I don't know' option was also included in all the perception tests in this dissertation (Figure 18).

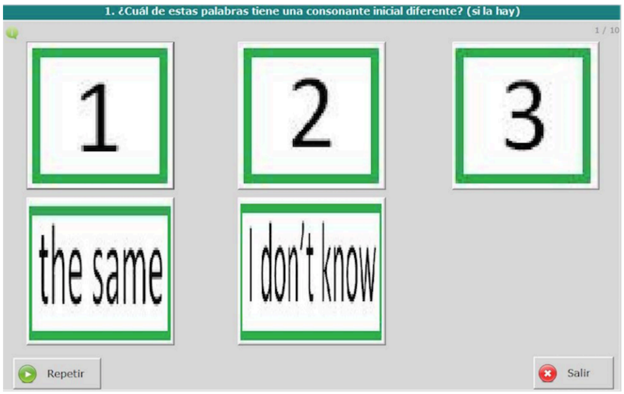


Figure 18. Sample screenshot of the discrimination tasks employed in studies 2 and 3

Following Flege and MacKay (2004), catch (or 'no-change') triads were included in order to test the learners' capacity to perceive different realisations of a particular phoneme as belonging to the same category. Flege and MacKay argue that as new phonetic categories are developed, sensitivity to differences between members of the new category and members belonging to other categories increases, but sensitivity to differences among members of that new category decreases. In this regard, stimuli in every triad were always pronounced by different male and female native speakers. Hence, even though triads such as *bag-bag-bag* could be judged as 'the same' (i.e. having the same phonological make-up), each stimulus was pronounced by a different speaker. This tested the learners' capacity to perceive different realisations of the same phoneme as belonging to the same category, disregarding acoustic features that are phonetically irrelevant to the identity of the sound (such as the speaker's voice or the phonetic context). Learners were allowed to listen to each triad twice. This should yield accurate representations of participants' phonological encoding, as they have several opportunities to make their choices, therefore reducing guesses (Baker & Trofimovich, 2006).

As for the identification tasks employed, they followed the conventional format. Learners were presented one stimulus at a time and they had to identify the sound they were hearing among a range of options. The options for the /s – z/ test were the two target sounds, the distractor /ʃ/, and the 'I don't know' button. The options for /æ ʌ ɑː ə/ included all the target vowels, /e/ as distractor, and the 'I don't know' button (Figure 19).

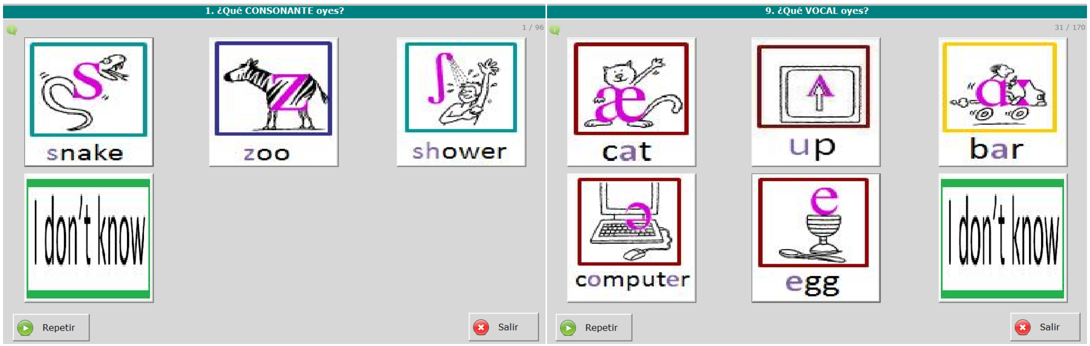


Figure 19. Screenshot of the identification tasks used in studies 2 (on the left) and 3 (on the right)

These two tasks were considered appropriate for the /s – z/ contrast and for /æ ʌ ɑː ə/ because the main problem Spanish learners have with these sounds has to do with the absence of these vowels and of the /s – z/ contrast in the phonemic inventory of their native language – therefore often resulting in the assimilation of these sounds into the Spanish phonetic categories for their closest counterparts due to equivalence classification (see Flege, 1987). However, measuring perception of /b d g/ by inter-language comparisons was not considered appropriate given that the differences between these English sounds and their Spanish counterparts are phonetic, rather than phonemic. That is, learners should not necessarily consider English and Spanish /b d g/ as 'different' sounds. Hence, comparing the acoustic characteristics of English and Spanish /b d g/ using English and Spanish words was considered problematic. Learners may have automatically marked stimuli as 'different' simply because they knew that stimuli featured two different languages, or perhaps based their choices on aspects other than occlusion (e.g. differences in VOT or neighbouring sounds).[75]

In light of the above, two different tasks were used in order measure the learners' perception of English /b d g/. The first was a delayed accent-mimicry task in which learners had to imitate the way English speakers pronounce Spanish. This was meant to test participants' awareness of the realisation of /b d g/ as stops in intervocalic positions. Although this task was aimed at measuring perception, it was part of the individual interviews for the production test, just after the picture-description task. In the second task, learners were asked to read a series of words to themselves in Spanish while listening to their English counterparts. They had to decide whether the degree of occlusion in the

[75] For example, in a triad including *ban (Eng.)* - *ban (Eng.)* - *van (Sp.)*, always marking the Spanish *van* (third-person plural of the verb *go*) as different just because of being Spanish. Additionally, in this type of comparisons with near-minimally paired words (English *ban* /bæn/ and the Spanish *van* /ban/), learners may focus on the vowel instead of on the target consonant, given that despite phonetic differences between English and Spanish /b/, the vowel is the sound that is actually different.

pronunciation of the target sounds they were hearing in the audio was the same as the one in the Spanish word that they had to read to themselves. This task was administered just after the identification task for /s – z/, in the same group interview.

In the delayed mimicry task learners imitated the Spanish accented speech produced by English native speakers when speaking Spanish, not by immediate imitation of a model presented, but recalling their impressions from memory (see Flege & Hammond, 1982; Rochdi & Mora, 2012). Accurate articulation of these non-distinctive phonetic differences between English and Spanish would suggest some kind of awareness of such differences on the part of the learners (Flege & Hammond, 1982), but also their ability to articulate these sounds correctly as stops in intervocalic position – which should not pose a problem for Spanish speakers given that /b d g/ are also pronounced as stops in Spanish in certain positions (see section 2.4.1 above). Following Flege and Hammond (1982) and Rochdi and Mora (2012), no demonstrations of English-accented Spanish were given, nor were learners offered instructions on the aspects of foreign-accented speech they had to concentrate on. As noted above, this was part of the production test, therefore avoiding possible training effects of the perceptual task that was to follow.

In the second task, learners had to decide whether the degree of occlusion produced in the articulation of a number of sounds was the same in Spanish and English. The researcher compiled a list of 48 English and Spanish cognates containing examples of voiced and voiceless stops in different positions. Even though the phonological structure of the cognates was not identical in terms of the neighbouring phonemes (e.g. having a different vowel sound after /b d g/ in some instances – the Spanish *abadía* [aβa̠ˈðia] vs. the English *abbey* [ˈæbi̠]), the criterion was to include pairs of words that would be perceived by students as related and/or similar. Participants were given a list of Spanish words with the orthography corresponding to the sound under investigation underlined and in bold. Next, recordings of the English versions of those words were played through individual headphones. Learners listened to one word at a time and, by reading the word in Spanish and pronouncing it to themselves, they had to decide whether the degree of occlusion in the pronunciation of the target sounds was exactly the same in English and Spanish (see section 4.2.2 for details).

It should be noted that the discrimination tasks for studies 2 and 3 were also administered to three native speakers in order to test the suitability of the stimuli. There were only two instances in which they did not discriminate the triads correctly, but these were reanalysed by the researcher and replaced with more suitable versions. The identification task was not administered to the native speakers, as they were linguistically naïve and were not familiar with English phonemic symbols. However, it was administered to the three expert judges that evaluated stimuli from the production tests (see section 3.3), as they were familiar with English phonetic symbols and were a reliable measure of the validity of the stimuli. There were a couple of incorrect identifications with two of the judges but after reanalysing the stimuli together, they all agreed that stimuli were appropriate.

3.1.3 Production tasks

a) Studies 1 and 2

The production tests in studies 1 and 2 were divided into three different tasks. The first was an imitation task aimed at measuring the learners' capacity to articulate the target sounds adequately after hearing them. This should provide a comprehensive picture of the learners' productive abilities, as it measured the learners' controlled, spontaneous and imitative production (comparable to the three modes of pronunciation practice put forward by Morley, 1991, 1994).[76] The second was a sentence-reading task intended to elicit the learners' pronunciation of the target sounds in a controlled manner (i.e. learners were supposed to have time to monitor their pronunciation). Finally, the third was a timed picture-description task aimed at eliciting a more spontaneous sample of the learners' speech, as they did not have much time to think before pronouncing the targets and they were also required to concentrate on meaning.

These three tasks provide a comprehensive picture of the learners' mastery of the target sounds, which is complemented by the information obtained from the perceptual tasks described above. As noted above, mastering a new phonological system involves mastering perception and production of the FL sounds and patterns. The two perceptual tasks for each set of targets offer a measure of the learners' perception of the target sounds (in theory, guiding their productive capacities – see section 1.1.5). The three production tasks reflect their ability to pronounce these sounds correctly, from imitative skills reflecting perceptual and articulatory ability, to controlled production in the sentence-reading tasks, and a more spontaneous pronunciation in the timed-picture description task.

---

[76] As noted above, the accent-imitation task aimed at measuring the learners' perception and production of English /b d g/ as stops was also administered in this production test, just after the picture-description task. However, it is not considered as part of the production test because it was intended to measure perception.

A great deal of pronunciation acquisition studies have used imitation tasks in order to measure the formation of new phonetic categories for FL sounds (e.g. Flege & Eefting, 1988; Flege & Hammond, 1982; Thomson, 2011). The results by Flege and Eefting (1988) suggest that ability to adequately imitate FL stimuli is linked with FL category formation. These researchers claim that learners' ability to imitate the target sounds adequately results from their capacity to perceive (i.e. categorise) these sounds correctly.

The imitation task was devoted to assessing the learners' capacity to attain the target sounds in production, regardless of their awareness of the phonological composition of words. An imitation task measures the participants' capacity to perceive and produce the target sounds without imposing any additional demands. It does not require them to be familiar with spellings, nor to be able to articulate the target sounds spontaneously, but simply to imitate what they hear. The imitation task could be considered as the 'easiest' (or less demanding) of the three tasks. Hence, speakers' mastery of a target sound should be represented by their ability to imitate it, followed by their ability to pronounce it in controlled production, and finally, in spontaneous speech.

The problem with imitation tasks is that they are not suitable for determining whether failure to imitate a particular sound is due to the subjects' perceptual or articulatory difficulties (Beddor & Gottfried, 1995). Furthermore, learners may be perfectly able to imitate a sound or pattern accurately, but fail to come up with that sound in a certain word because of the spelling, or lack of familiarity with the word and its phonological composition. For example, using the adequate consonant for words like *Stephen* /v/, *Philip* /f/, *laugh* /f/, or *daughter* – silent. Additionally, even if learners are able to attain the sound in production and know where to pronounce it, they may fail to attain an adequate articulation when engaged in spontaneous conversation, be it due to the anxiety of the situation, speed or cognitive demands of the conversation, etc. Hence, in order to obtain a more comprehensive measure of the learners' pronunciation ability, two more tasks were included, namely a sentence-reading task and a timed picture-description task.

The sentence-reading task assesses the learners' capacity to pronounce the target sounds correctly, showing both whether they know what the pronunciation for those words is and also whether they can articulate those sounds adequately. That is, they do not only need to be able to articulate the sound correctly, but when they read the words, they also need to know *what* sound is the one they need to pronounce (e.g. knowing that *cat* is pronounced with /æ/ and not /ʌ/). This can be particularly difficult for sounds that are not always represented by the same orthography (e.g. /z/ in *president*, given that <s> can be pronounced with /z/, as in *music*; with /s/, as in *person*; with /ʃ/, as in *censure*; with /ʒ/, as in *leisure*; or be silent, as in *aisle*). Thus, words in this task were divided into familiar and novel words. Familiar words were intended to test improvements in words learners had seen in training (i.e. words whose phonological make-up should be familiar to learners). Novel words were aimed at measuring whether they could generalise improvements to words that they had not seen during training.

Finally, in order to measure the learners' spontaneous production of the target sounds, a timed picture-description task similar to the one used by Saito (2013) and Saito and Lyster (2012) was used. In this task, learners were presented with a number of pictures they had to describe, together with several words to guide their description, including target words and distractors. Learners were told to describe what they saw on each picture in the shortest time possible. They were allowed to create more than one sentence, but it had to be the first thing that came to their mind (see Figure 20). Despite offering words as prompts, it was considered that the timed nature of the task should be cognitively demanding enough so as to obtain a reliable sample of the learners' spontaneous pronunciation (Derwing, Rossiter, Munro, & Thomson, 2004; Saito, 2013), as learners find it difficult to concentrate on phonetic form when the task also requires them to focus on meaning (Guion & Pederson, 2007). The initial idea was to use a free narrative task that would prompt learners' genuinely spontaneous productions. Nevertheless, this was problematic for the purpose of the current studies as it would be difficult to elicit an equal number of examples of the target sounds from every participant – consequently making data comparisons difficult.

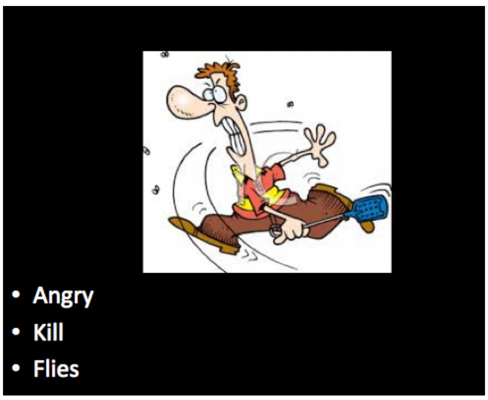


Figure 20. Sample prompt from the picture-description task used in study 2

It is important to point out that some studies measuring participants' pronunciation in FL acquisition research test the learners' production by eliciting several repetitions of the same word in order to obtain average scores for the pronunciation of a particular target sound in the same word at different times (e.g. Lord, 2010; Rochdi & Mora, 2012). In studies 1 and 2 in this dissertation, the input learners were going to receive was subject to the materials available and was consequently limited to a number of target words (as opposed to, for example, edited or modified stimuli, or non-sense words aimed at measuring certain sound contrast that are edited and created by the researcher). Thus, including a range of different words that served to assess the target sounds in different contexts was considered more appropriate than focusing on only one word pronounced several times. Additionally, this minimised the risk of learners noticing the words they were exposed to during training were the same words as those in the pre-tests and trained conscientiously to perform adequately in the post-test.

b) Study 3

Learners' improvements in study 3 were measured by a much simpler task than the ones used in the previous studies. As noted above, study 3 only measured the learners' pronunciation of the targets addressed, as including a perceptual test with the target stimuli could have training effects that would not allow a reliable measure of gains obtained from instruction.

Since the training stimuli in study 3 (i.e. the target words) had to appear in the pre- and post-tests in order to measure potential learning gains, an attempt was made to avoid learners conscientiously studying for the post-test. The aim was to ensure that possible improvements in the pronunciation of the target words were a result of instruction and not because of participants seeing (or looking up) the words somewhere else. Hence, the pre- and post-tests in study 3 consisted in an oral matching task in which participants had to associate words with colours based on their impressions. Learners were initially told that they were going to take part in a study where they would receive a number of tips about English, but that the researchers' goal was to investigate synaesthetic word-colour associations by FL learners. The idea was to offer them free tips in exchange for their participation in the researchers' synaesthesia project. The aim of this task was to elicit the participants' pronunciation of the target words without them knowing that pronunciation was the focus of the study.

Students were presented the testing stimuli in blocks of five in a PowerPoint slideshow presentation. Each slide in the PowerPoint included five items (target words and distractors) and three colours from which students could choose in their associations with the five items (Figure 21). Participants had to pronounce every word in the slide in the carrier sentence '_____ *is* _____', indicating the colour each word evoked (e.g. '*Aisle is green*').

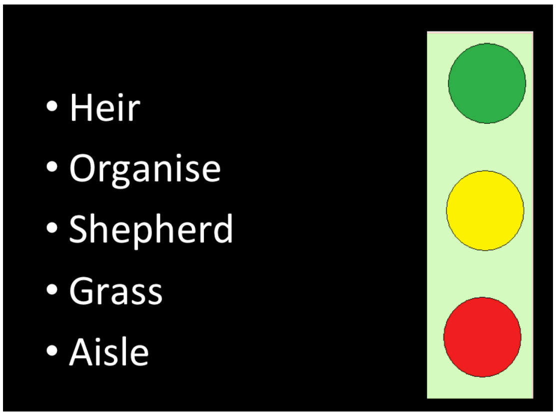


Figure 21. Sample slide from the pre- and post-tests for study 3

3.2 Testing procedure

Before the beginning of each study, participants were required to sign two documents. One was a consent form informing them of the purpose of the study, which included a confidentiality statement by the researcher whereby anonymity of students was guaranteed during the entire process and after it (see Appendix 1). The other was a confidentiality agreement by which students committed not to discuss any information regarding the study with anyone until the study was over, including other classmates (see Appendix 2). This last form was considered essential in order to avoid students in the experimental groups discussing the issues covered in training with classmates in the control group (or receiving a different type of instruction), which may affect the results obtained by the latter group in the post-test.

a) Studies 1 and 2

Even though perceptual tests are described first in order to follow the same structure outlined above, production tests were always administered first to avoid possible training effects. Perceptual tests were conducted in groups of a maximum of 17 students in a quiet computer room at university. They were administered using TP, an open-source application for designing and administering speech perception tasks (Rato, Rauber, Kluge, & Santos, 2015). The tests were run on Intel Pentium 4 computers (2.3 GHz, 1GB RAM, 80 Gb hard drive). Stimuli were played over RoyCan RC-150D headphones.

Even though the perceptual test was conducted in groups, participants took the test individually at their computers. Given the nature of this test, learners were given clear instructions to remain quiet during the test. They were instructed to remain silent in their seats during the whole session, even if they finished.

Perception tests lasted between 40 minutes and an hour (with a brief pause between the two tests – see the sections below for details). Apart from the pause between these two parts, there was a pause after every 20 stimuli in order to avoid fatigue. A pop-up screen with instructions in Spanish appeared every time an activity finished. This was done to remind participants of the focus of the next activity (e.g. vowels or consonants) and the target aspects they had to pay attention to (e.g. for /s – z/, whether they had to pay attention to sounds in initial, medial or final position – see sections below). In any case, instructions were always accessible from one of the buttons on the screen.

The discrimination task was administered first in order to avoid possible training effects. It was considered that learners might perform better in the identification task if it was conducted before the identification task. As noted above, discriminating whether the three stimuli presented are the same or different should be easier than labelling them.

Following Flege, MacKay, and Meador (1999), students were familiarised with the oddity discrimination task by means of a two-minute training session with geometric shapes. Learners were shown a PowerPoint presentation in which they had to identify the odd item out among three different geometric shapes. As in the actual discrimination task, both change and catch triads were included. Change triads were composed of two identical geometric shapes and one that was different (e.g. two squares and a circle), catch triads included three identical figures. Slides were shown one by one and participants had to say '1', '2', '3', or 'the same' out loud, depending on what they saw on the slide (see Figure 22). During this training session, immediate oral feedback was given after each slide.

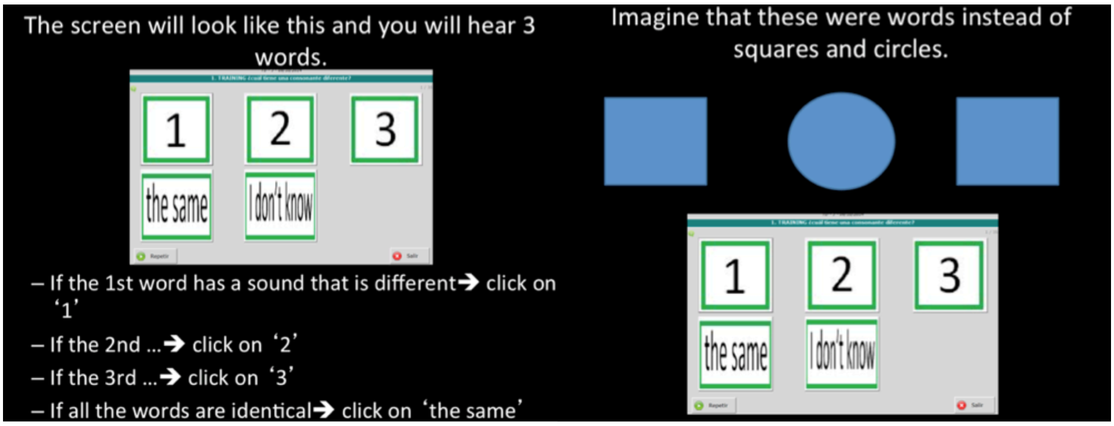


Figure 22. Training session for the oddity discrimination task (instructions on the left, visual oddity training on the right)

In addition, the first 10 triads in the auditory test were also intended as training for the oddity task. They were exactly like the actual test, although triads did not include any of the target sounds. This initial training included sounds that were phonetically very distant (e.g. *ping-pong-ping*) with the intention of familiarising participants with the aural version of the oddity task before starting the test.

The identification task followed the conventional format of this type of task. Words were presented individually and learners had to decide which sound they heard among three options, including the target sounds, a distractor, and the 'I don't know' option mentioned above.

One of the problems with identification tasks is that labels need to be provided for the different choices available, something that is not always easy as the choices may affect the listeners' performance (Logan & Pruitt, 1995). For study 2 this was not a problem, as the phonetic symbols for /s/ and /z/ are quite straightforward to interpret, as they also coincide with some of their most common spellings. Besides, learners were already familiar with the symbols thanks to the oral skills course they received during the previous year, and also due to the phonetics course on which they were enrolled at the time. However, the labels required for the target vowel sounds in study 3 were more problematic, as the phonemic symbols and their orthographic representations do not coincide and this may be a confounding factor for participants. Following Carlet and Cebrian (2014), each label (i.e. phonemic symbol) was accompanied by a keyword that consisted of a high frequency sample word featuring that sound (e.g. /ɑː/ as in *bar*, /ʌ/ as in *up*, /æ/ as in *cat*).

With regard to the production tasks, they were conducted individually with the researcher in a quiet room at university. The tests lasted around 10 minutes. They were administered as a PowerPoint presentation on a MacBook Pro computer. Tests were recorded with a SAMSON C01U USB Studio Condenser Microphone (16-Bit resolution, 44.1-48kHz sampling, large 19mm diaphragm microphone) using Audacity (The Audacity Team, 2015). The microphone was placed at a distance of approximately 30 cm from the speaker's mouth, on a cylindrical support made of cardboard. Learners controlled the PowerPoint so that they could pass the slides at their own pace. At the beginning of each task, a slide with instructions was presented in Spanish in order to make sure that they understood the instructions for each task (Figure 23). The instructions informed them that they were going to see one sentence per slide. They were told that they should read each slide carefully before saying them out loud, trying to pronounce the sentences the best they could.

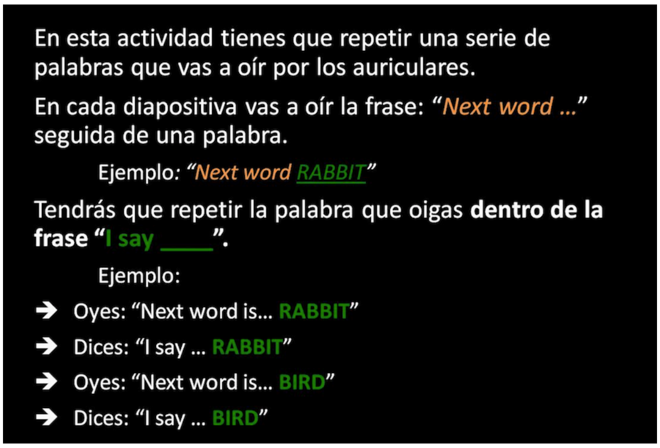


Figure 23. Sample screenshot of the instructions learners saw at the beginning of each task

In order to minimise anxiety, a factor conditioning language acquisition in general (Krashen, 1985) and pronunciation performance in particular (Baran-Łucarz, 2014; Guiora et al., 1972, Guiora et al., 1980; Schumann et al., 1978), learners were reminded that the recordings were anonymous and informed that the researcher would not listen to them as they performed the test. He would be in the same room as students in order to clarify possible questions, but he would be listening to music using noise-cancelling headphones. This was considered essential in order to obtain a realistic, anxiety-free sample of their pronunciation, as the fact that they were being recorded could add even more anxiety to the already stressful situation of being tested.

The first task was the sentence-reading task, followed by the timed picture-description task, and finally, the imitation task. The imitation task was administered last in order to avoid possible training effects. In study 1, the delayed accent-mimicry task for /b d g/ was conducted just after the timed picture-description task.

For the sentence-reading task, following Flege and Hammond (1982), test items were presented in the carrier sentence *'The ____ is on/in the ____'*. This allowed controlling the surrounding phonetic environment in which target sounds appeared. The target items were pronounced in a stressed position in the sentence, which increased the likelihood of learners articulating the segments more clearly. For /b d g/, the frame was slightly modified in some sentences given the need to include the target words in the minimum number of sentences possible due to time constraints (an attempt was made to include at least two target words per sentence). However, test items were always presented in stressed positions and, for items addressing /b d g/, flanked by vowels. It must be pointed out that the first three sentences were common sentences such as *my name is John* or *I love cinema* intended to help students get familiarised with the procedure. These were not taken into account for the analysis.

The imitation task was administered last in order to avoid training effects from the other two tasks. The models for imitation were obtained from the pronunciation dictionaries mentioned above. Following the elicitation technique used by Thomson (2011), target words were edited and embedded in the carrier sentence '*Next word* _____', recorded by a female English native-speaker from Brighton (UK). The target word was played approximately 0.7 seconds after the prompt by the native speaker. This sequence was played for each target word, to which participants had to reply '*I say* _____', imitating the model pronunciation, in this case, always obtained from the dictionaries above. No written representations of the target words were offered. This should force learners to rely exclusively on their perception, therefore avoiding possible interference with spelling. Learners only saw the written prompt *'Next word ____'*, and the target word was automatically played as they moved to the next slide of the PowerPoint. The target words for imitation were played over a pair of Sony MDRZX100 ZX headphones.

b) Study 3

The production tests for this study were conducted in a quiet office at the University of Murcia. The office was divided into two big rooms, one was the room where participants completed the interviews and the other was the researcher's office. The researcher was sitting in one of the offices with the door connecting the two rooms open while participants completed the interviews. Interviews were recorded with the built-in microphone in the same MacBook Pro computer where the PowerPoint was presented. Each interview lasted approximately 5 minutes.

3.3 Evaluation procedure for the production tasks: Non-native expert judges

The evaluation of participants' productions in studies in this dissertation was made by non-native judges expert on English pronunciation. A great deal of studies have used native-speaker judges in order to assess learners' oral productions (e.g. Anderson-Hsieh et al., 1992; Derwing et al., 1998; Derwing & Rossiter, 2003; Tanner & Landon, 2009), some have used a combination of native and non-native judges (e.g. Ducate & Lomicka, 2009; Lord, 2008), and some have used exclusively expert non-native judges (e.g. Zampini, 1996).

Native speaker judges offer appropriate measures of aspects such as intelligibility, comprehensibility or native-likeness (e.g. Anderson-Hsieh et al., 1992; Derwing et al. 1998; Derwing & Rossiter, 2003). However, the evaluations performed by native judges are often based on an overall impression of pronunciation ability (Anderson-Hsieh et al., 1992). For more precise judgements requiring special attention to subtle phonetic details, native speakers with no training in phonetics and phonology may fail to provide adequate ratings given that they are often unaware of how the phonology of their native language works and they may incorrectly judge something as correct as long as it is intelligible. Native judges[77] could provide reliable measures of accuracy if, for example, minimal pair words were used to test the appropriateness of learners' productions. For example, rating vowel correctness by deciding which of the three words they hear *buck-back-bark* (see e.g. tasks used in Flege, MacKay, & Meador, 1999). Nonetheless, when stimuli are not grouped into minimal pairs, as is the case of the studies in this dissertation (see explanation below), native speakers may unconsciously rate the production of specific items as correct because they consider it acceptable (i.e. not affecting intelligibility). Native judges may have the same problem in study 4, possibly rating mispronunciation such as *fam<u>ou</u>s* with /ʌ/ instead of /ə/ as correct as long as it is intelligible for them.

With regard to the number of judges employed, they vary depending on the aspect under evaluation. Some studies addressing intelligibility, comprehensibility, or accentedness have used up to 48 listeners in order to provide a reliable measure of the aspects affecting those constructs (e.g. Derwing et al., 1998). However, in same study, Derwing and

[77] It is assumed that suitability of judges for stimuli in this dissertation depends on them being expert on English phonetics and phonology, not on their being native speakers of the language. Due to a lack of availability of expert native judges, three non-native judges expert on English phonetics and phonology were used. For the sake of simplicity, when the author mentions 'native judges', he refers to naïve native speakers (i.e. not experts on English phonetics and phonology) in order to establish comparisons between expert judges (native or non-native) and non-expert native speaker judges.

colleagues used only 6 judges in order to rate speakers' fluency. Similarly, Derwing and Rossiter (2003) used 6 expert teachers in order to rate comprehensibility, accentedness and fluency. Tanner and Landon (2009) used 10 non-experienced listeners to rate comprehensibility, while Ducate and Lomicka (2009) used two. Nonetheless, studies using expert judges often make recourse to a smaller number of listeners. For example, Anderson-Hsieh et al. (1992) used three, Saito and Lyster (2012) and Thomson (2011) employed 5, and Akahane-Yamada et al. (1997) only one.

The expert judges used in dissertation are three non-native judges expert on English pronunciation.[78] They have all followed BA courses on English pronunciation, English phonetics and phonology, and English intonation, in addition to an MA course in acoustic phonetics with a focus on English. One of them holds a PhD in English Linguistics (more specifically, in sound-symbolism) and the other two are currently completing their PhDs on the acquisition/learning of pronunciation with new technologies (one is the researcher). They are experienced EFL teachers and their native language is Spanish (for judges 1 and 2) or Spanish-German (for judge 3). Hence, they are very familiar with the interlanguage of Spanish learners of English. In addition, a fourth expert judge was used in order to disambiguate disagreements in the other three judges' decisions. This judge was also a native speaker of Spanish with an MA and a PhD in English Phonetics and ample experience teaching English Phonetics and Phonology.

The fact that the researcher was involved in the evaluation of the data should not compromise the results as long as stimuli were anonymous and randomised. In fact many researchers often act as raters in the evaluation of their data (see e.g. Derwing & Rossiter, 2003; Thomson, 2011; Zampini, 1996). In this respect, participants in the studies reported in this dissertation were always anonymous, as they were assigned a participant number at the beginning of the studies. Moreover, recordings were always saved with the participant number, rather than the participant's name. Furthermore, participant numbers were not assigned in order (i.e. per group, with Group 1 having numbers 1-20 and Group 2 from 20-40). Instead, they were distributed randomly, with numbers as distant as 1 and 100 possibly belonging to the same group. Thus, the researcher did not know whether participants had received training or not during the evaluation of stimuli.

---

[78] Given the simplicity in the evaluations of study 3 as compared with the judgements required for studies 1 and 2 (given their subtle phonetic nuances required for vowel sounds, for example), only two judges were used in study 3. The learners' mistakes in study 3 were easily noticeable, for example, substituting the /k/ in archives with a /tʃ/, or pronouncing an /s/ in *aisle*).

Some studies use identification scores assessing whether listeners correctly identify the intended target sound (e.g. Flege et al., 1999), others employ rating-scales measuring the degree of intelligibility, comprehensibility, accentedness, or any other construct (e.g. Derwing & Munro, 1997; Ducate & Lomicka, 2009; Tanner & Landon, 2009), and others combine the two measures (e.g. Akahane-Yamada et al., 1998). In this dissertation, production measures are always quantified as either correct or incorrect. Each token was assigned a value of either 1 or 0 (1 if the target sound pronounced correctly, 0 if the sound was mispronounced).

The original intention was to use TP for the evaluation of stimuli. Even though it was originally devised for speech perception tasks, it can be easily adapted for stimuli rating by introducing items as though it were a two-forced choice identification task in which raters mark stimuli as being either 'correct' or 'incorrect' (or 1/0). This would allow judges to perform evaluations directly from their computer and at their own pace. Nevertheless, given the programme's original purpose, introducing stimuli for rating was rather inconvenient, as each stimulus had to be introduced individually and a random 'correct' answer had to be assigned for every stimulus. As a case in point, in study 2 there were a total of 123 target words, and introducing stimuli took around 10 minutes per word (123 target words x 45 participants x 2 testing times). Thus, this slowed the process considerably. Finally, TP was discarded for rating the stimuli in studies 2 and 3 due to technical and time limitations. After having introduced the 10945 tokens for study 2 (approximately 20 hours), the programme could not handle such a large amount of data and the test did not work.

Rating sessions were conducted in a sound-attenuated professional studio at UCAM University. Given the high number of tokens for evaluation in the three studies in this dissertation, listening sessions were conducted over a period of three months to avoid listener fatigue. Rating sessions were often 1h 30 minutes long, depending on the availability of the judges and the studio. Judgements were made individually, but all the judges were present in the listening sessions. The researcher randomised the stimuli and played it with a MacBook Pro computer using a Belkin headphone audio splitter that outputted sound over three different SONY MDRZX100 ZX headphones. Each stimulus was played as many times as judges needed. Recordings were played using only one computer so that sound quality was the same for every judge. However, the evaluations from each judge were annotated on three different laptops that judges handled individually. Judges made their evaluations on their own laptop on an Excel spreadsheet.

At the beginning of the listening sessions for each sound, a brief training session was given on how to rate that particular sound. The researcher played a number of sample words containing the target sound from several online dictionaries and then played a number of random recordings (without judges knowing which) in order to discuss together what ratings those recordings would receive. Given that listening sessions for each sound were always conducted over different days, a brief round of training was made at the beginning of every listening session.

All the stimuli for each target sound were evaluated together, task by task (e.g. starting with stimuli from the sentence-reading task for /z/, followed by the ones from the timed picture-description task, and finally, from the imitation task). This was a way of accustoming listeners to rating a particular sound, therefore acting as some kind of training too. Once the evaluation for a sound was completed, the researcher compared the judges' scores and marked disagreements in a different colour in order to re-judge them, send them to a fourth judge for disambiguation, or analyse them using spectrographic analysis. Re-evaluations were made on a separate spreadsheet; the original evaluations by each judge were not altered in order to obtain inter-rater reliability measures.

Additionally, intra-rater reliability was calculated for each judge by playing a random selection of 20 extra items that had already been assessed, in order to test whether judges were consistent in their evaluations. An equal number of items for the intra-rater evaluations were obtained from pre- and post-tests, and an equal number of items were used for each sound under evaluation.

The 10-minute recordings from the production pre- and post-tests were edited individually. Each recording was normalised for pitch amplitude and background noise was reduced (noise reduction 24dB, sensibility 0dB, smoothing 150Hz, 0.15s) using Audacity. Each token was saved as a separate .wav file and labelled as the target word it contained, indicating the participant number and whether it belonged to the pre- or post-test (e.g. *fuzzy_4_pre*). Recordings for each task (imitation, sentence-reading and timed-picture description) were saved into different folders.

The evaluations in these studies were categorical (acceptable/non-acceptable), based on the judges' impressions on whether sounds were articulated correctly. Using judges for evaluation has several advantages, such as the fact that decisions are based on the opinion of three different experts, or that it is comparatively faster than analysing each token individually with spectrographic analysis. Subtle improvements from pre- to post-test may have gone unnoticed given that stimuli have not been measured acoustically. Nevertheless, the judges' decisions should reflect clear improvements in the learners' productions from pre-to post-tests, as very subtle nuances in acoustic measurements may not even be noticeable for human raters.

As pointed out above, studies in which stimuli are evaluated by human judges commonly employ identification tasks in which listeners hear the word pronounced by the student and they have to decide among minimally paired words (e.g. Flege et al., 1999). Identification tasks in which judges have to choose among English minimally paired words were discarded in this dissertation given that Spanish learners' interlanguage productions do not always adhere to the vowels in minimally paired words. As a case in point, assigning Spaniards' productions of /æ/, /ʌ/ and /ɑː/ to one of the three possible options in *hat-hut-heart* would be problematic given that Spanish EFL learners tend to pronounce these English sounds as the /a/ vowel in the phonemic inventory of Spanish. Thus, an identification task with three English words featuring the three different English sounds was not considered adequate to detect mispronunciations such as *[hat] (with a Spanish /a/) for either *hat* or *hut*, as listeners may incorrectly judge mispronunciations as correct if they identify them with one of the given options.

For studies 1 and 3, evaluations of stimuli were made word by word, rating all tokens of a given word by every participant, first one testing time (pre- or post-test), then the other. As a case in point, judges rated all tokens from the pre-test for the word *above* from every participant; then, they rated all the tokens from the post-test. The researcher randomised which testing time would be presented first for each target to avoid that ratings always started with either the pre- or the post-test productions. Even though the researcher was aware of this arrangement (as it was him who was playing the recordings), the other two judges did not know whether stimuli belonged to the pre- or post-tests. In addition, as noted above, despite the fact that the researcher knew whether recordings belonged to pre- or post-test, he was not aware of whether the participant under evaluation had received training on a particular sound or not, as participants were randomly assigned to their groups.

The evaluations for study 2 were conducted in a slightly different manner. The approach adopted was similar to the one employed by Akahane-Yamada et al. (1997). These researchers played randomised productions from pre- and post-tests and asked judges to decide which one was better. In this study, pre- and post-test productions were also compared, but judges based their comparisons on whether productions were adequate realisations of the target sound or not. This was done in order to facilitate judgements by direct comparison between the two productions. Judges had to identify whether learners 'pronounced the right sound' in the target words before and after receiving instruction. It was some kind of identification task, not focusing on which word was pronounced, but on whether the target sound was articulated adequately or not (i.e. *do you hear x sound in this word? Yes/no*). Judges marked each stimulus with either 1 or 0 depending on whether learners attained the sound in production, 1 if production was accurate, 0 if it was not. Nevertheless, this direct comparison between pre- and post-test productions allowed the researchers to signal when one of the two productions was markedly better than the other. That is, if learners already pronounced something acceptable in the pre-test but improved in the post-test (i.e. receiving a score of 1-2), or if they pronounced something acceptably in the pre-test but failed to pronounce it adequately in the post-test (i.e. 2-1).

Ideally, pre- and post-test stimuli should have been randomised, but because of the technical problems encountered and the need to make the evaluations manually, pre-test productions were always presented first. Since judges were writing down their answers on an Excel spreadsheet, they needed to know where to write their marks as stimuli were played. This was done for study 1 by always presenting first the pre- or post-test for a whole set of tokens (i.e. all participants pre-test productions of *fuzzy* and then all the post-tests productions). This allowed the researcher to mark which testing time was presented first and then organise the corresponding Excel columns after the evaluation. However, since in study 2 judges were directly comparing pre- and post-test productions, it would have been very complicated to change the order of presentation for every participant (e.g. participant 1 (pre); participant 2 (post); participant 3 (post), etc.). Nevertheless, like in study 1, judges did not know which participants belonged to the experimental and control groups, as participant numbers were randomised across groups.

Chapter IV. Exploring podcasts' potential for pronunciation training

This chapter deals with the simplest of the three technologies explored in this dissertation: podcasts. They differ from similar tools, such as cassette tapes, or CD-ROMs, in that they do not only offer a virtually unlimited source of input, but also allow learners to record their output and share it with other people. The chapter addresses podcasts' potential to improve the learners' perception and production of a number of target sounds that are considered challenging for FL learners. The approach adopted in these two studies exploits the potential of podcasts for three different aspects: input, output and peer evaluation.

The chapter describes two studies. The first is a pilot study that explores this technology for pronunciation training in a non-controlled environment. Instruction was offered entirely online, therefore avoiding the laboratory-like environments sometimes used when testing certain CALL tools. The second is a modified version of that pilot study that attempts to overcome the limitations encountered in that preliminary experience. Situdy 1 entailed a closer supervision by the researcher, but the findings obtained offer valuable insights for the implementation of this tool for autonomous learning.

4.1 A pilot study

4.1.1 Aims

Given the complexity of the three-stage weekly procedure students had to follow in the podcast-based instruction, and because of the technological demands imposed by the different tasks, such as downloading podcasts and recording their own, uploading them onto an online learning platform, or evaluating their peers' recordings through Google Drive, a pilot study was designed in order to test the method and the feasibility of the tasks.

The data obtained from the perception and production tests in this pilot study have not been analysed, as the aspects addressed here were not the pronunciation targets intended for this dissertation. Nevertheless, the participants' responses to the questionnaires and the information obtained in this preliminary study offer valuable insights into the use of podcasts for pronunciation training, serving not only as scaffold for the design of study 1, but also as a guide for practitioners who wish to implement podcasts in their classes. Thus, before reporting on the data obtained in study 1, an account of the procedure followed in the pilot study and a summary of the findings that led to the final design of study 1 is provided.

4.1.2 Method

a) Participants

Participants were recruited from an oral skills course in the first year of the degree in English studies at the University of Murcia. The study was advertised as a free five-week pronunciation course that would complement the information received in the course they were enrolled in. They offered to participate in the study as something extra and completely voluntary that would take place outside their classroom. Participants were offered 0.3 points of their final mark in exchange for their participation in the study. Seventy participants contacted the researcher (not the students' lecturer) showing their interest to take part in the study. However, only 58 completed the two initial interviews and started completing the weekly tasks.

The initial 58 participants were randomly assigned to two groups, control (n = 28) and experimental (n = 30). Thirteen participants in the experimental group abandonded the study during the first two weeks. They claimed to be very stressed with their many university commitments and argued that they would not be able to follow the three-step procedure they had to complete every week. Thus, only 45 participants finished the study. The final groups were slightly unbalanced, with 27 students in the control group and 18 in the experimental group (N = 45). These were 10 male and 35 female (mean age 18.8; SD = 1.21).

b) Instruments

Three main tools were used in this pilot study: 'BBC 6 minute English' podcasts, Edmodo, and Google Drive online questionnaires. Podcasts were the core tool of the study, as they are one of the tools under evaluation in this dissertation. Because the model employed in the oral skills course in which students were enrolled was British English, all podcasts were obtained from the 6-minute English series offered by the BBC.[79] As the name indicates, this series of podcasts offers weekly recordings of 6 minutes on a wide range of topics. A further reason for the selection of this series is the broadcasters' accent, considered to be SSBE, the model taught to the students in their course.

The second tool used was Edmodo,[80] which served as a private educational online platform through which to share materials and interact with students. Given that participants were not the researcher's students, Edmodo proved to be a convenient alternative to the virtual campus offered by the university. This platform allows teachers to create several online groups and invite as many students as needed. Besides, the private nature of this tool allowed the teacher to restrict access to the group (and consequently to the training materials), which might have not been as convenient with the university virtual platform, as the control group would have also had access to the training materials. Furthermore, this also guaranteed anonymity of participants, given that they were encouraged to create a nickname to interact on the platform – something also problematic with the university's virtual learning environment (VLE), which shows students' name and surname(s) by default. In this way, students were able to upload their recordings anonymously (especially convenient for shy students), which were only visible for students participating in the project.

Finally, Google Drive's online questionnaires were used to gather anonymous data for the peer and self-evaluation of students' recordings. A new questionnaire was created each weak in order to assess the aspects covered in that week's podcast. This enabled students to make evaluations comfortably from their homes, also allowing the researcher to monitor participation and access the evaluations as soon as they were completed. Furthermore, Google saves the responses in an Excel spreadsheet that allows for a fast and convenient handling of data.

c) Training stimuli

Since the study was offered as a free pronunciation course, training was designed with the intention of offering as wide a selection of aspects as possible. This selection was based on

---

[79] BBC 6 minute English: http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/

[80] Edmodo is an online educational platform similar to a social networking service that is freely available on the Internet. It allows users to engage in discussions on a private forum, sharing files to a closed group of users, etc. It can be used as desktop tool but also as an application for smartphones and tablets. https://www.edmodo.com/?language=es

the syllabus of a phonetics course in the second year of the same degree, focusing on issues notoriously problematic for Spanish speakers. Despite the fact that some of these aspects may not always be detrimental to the learners' intelligibility, they were considered important for learners aiming at high levels of oral proficiency (see note 67 above).

In light of the above, training was planned for five weeks that dealt with the following:

1. The pronunciation of English stop plosives (/p t k b d g/) and the /b – v/ distinction.
2. The /ɪ – iː/distinction.
3. Word and sentence stress, contrastive stress, and the importance of /ə/ for English rhythm.
4. Weak forms and revision of schwa.
5. The vowels /æ/, /ʌ/, and /ɑː/.
6. Falling intonation in wh-questions.

d) Training procedure

This study was conducted during the second semester of the academic year 2013/2014 (March-May). At the beginning of the study, students were assigned a participant number in order to protect their anonymity. Participants in the experimental group were informed that training would last over a period of five weeks and that they had to complete four interviews, two at the beginning of the study and two at the end. Participants in the control group were told that given the high number of volunteers, it was not possible to offer training to the whole group at once, but that they would continue to receive the 0.3 points of their final mark provided that they completed the four interviews. They were informed that they would have the opportunity to receive instruction during the following year, as part of a phonetics course in the second year of their degree.

Numerous researchers have emphasised the need to move beyond laboratory-like, controlled environments in CAPT (Lord, 2010; Olson, 2014; Wang & Munro, 2004). Thus, an effort was made to make this study as unrestricted as possible, emulating the way learners could use this technology as a complement to classroom instruction. Hence, everything was done from home, except for the pre- and post-tests.

The weekly procedure students had to complete was as follows (see Figure 24 for a summary of the stages described below). First, the researcher uploaded a podcast and a PowerPoint presentation onto Edmodo. The podcast was the input learners were going to receive for that week and the PowerPoint included information about the peculiarities of the problematic aspects for that week together with tips on how to correct possible mispronunciations. The first step was to read the presentation prepared by the researcher in order to learn about the problematic aspects for that week and then listen to the podcast carefully, paying special attention to the features pointed out by the researcher and concentrating on the way those features were pronounced by native speakers.

Stage 1. Input stage

- Theory presentation
- Listening to the podcast

Stage 2. Output stage

- Recording a podcast

Stage 3. Podcasts evaluation

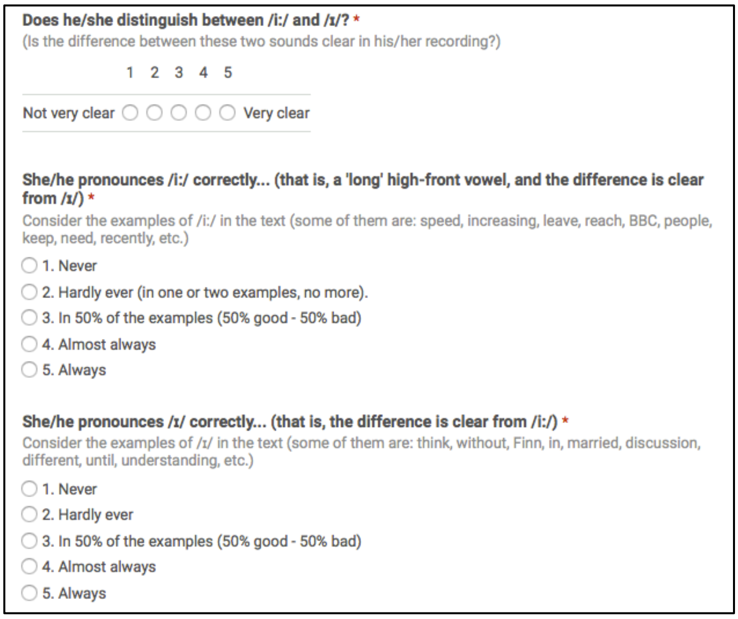
- Self-evaluation
- Peer evaluation

Figure 24. Stages learners had to follow weekly in the pilot study

At the second stage, participants had to record their own version of an excerpt from the same podcast and upload it onto Edmodo. This was meant to elicit samples of the target features, offering learners production practice for those key features. A detailed set of instructions was uploaded every week, including the text that they had to record for that week and indications on how to perform these stages step-by-step (see Appendix 3 for an example). Furthermore, they were asked to give their opinion at the end of their recording. Thus, two goals were met. On the one hand, this was an attempt to foster meaningful listening at the input stage. On the other hand, this was intended to elicit a sample of a more spontaneous production.

An additional set of instructions was created in order to inform participants of different ways in which they could record their podcasts. Participants were encouraged to record their podcasts using Audacity on their computer, given the free accessibility of this audio software and the simplicity of converting files into mp3 format. However, they were also offered the possibility to use Podomatic,[81] or any smartphone app that allowed them to make the recording from their phone[82] and upload them directly onto Edmodo using the Edmodo app.

Finally, the third stage consisted in the self-evaluation of the pronunciation of the target aspects, together with the peer-evaluation of two of their classmates. Thus, the researcher attached a list with two random candidates assigned to each participant in the same post that contained each week's instructions. The aim was that all participants should receive peer-evaluations from two different people, with these being reassigned and randomised every week. Both the peer-evaluation and the self-evaluation were carried out through Google Drive's questionnaires, with clear guidelines about what to evaluate and sample words that students could pay attention to for the assessment of those aspects (see Figure 25 for an example).



Does he/she distinguish between /i:/ and /ɪ/? *
(Is the difference between these two sounds clear in his/her recording?)

1 2 3 4 5

Not very clear ○ ○ ○ ○ ○ Very clear

She/he pronounces /i:/ correctly... (that is, a 'long' high-front vowel, and the difference is clear from /ɪ/) *
Consider the examples of /i:/ in the text (some of them are: speed, increasing, leave, reach, BBC, people, keep, need, recently, etc.)

○ 1. Never
○ 2. Hardly ever (in one or two examples, no more).
○ 3. In 50% of the examples (50% good - 50% bad)
○ 4. Almost always
○ 5. Always

She/he pronounces /ɪ/ correctly... (that is, the difference is clear from /i:/) *
Consider the examples of /ɪ/ in the text (some of them are: think, without, Finn, in, married, discussion, different, until, understanding, etc.)

○ 1. Never
○ 2. Hardly ever
○ 3. In 50% of the examples (50% good - 50% bad)
○ 4. Almost always
○ 5. Always

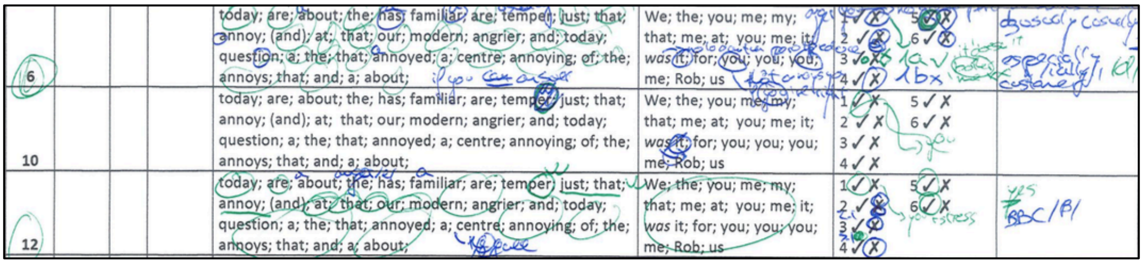
Figure 25. Sample of the peer evaluation questionnaire for /ɪ/ and /i:/ in the pilot study

The peer-evaluation questionnaires addressed general and specific aspects of the learners' pronunciation. Peers were required to rate the recording using Likert-type scales to evaluate overall pronunciation ability on aspects such as degree of foreign accent, degree of comprehensibility, or liveliness of the recordings. Nonetheless, they were also asked to focus on the key aspects that were the goal of the study. After participants completed the evaluations for their two peers on Google Drive, they were required to post a summary of

[81] Podomatic is a website that allows users to create and upload their own podcasts without the need to download additional software: https://www.podomatic.com/login
[82] The one recommended by the researcher was the *Voice Record App* (Dayana Networks Ltd., 2015).

their assessment under the participant's podcast in Edmodo so that (s)he would know what aspects could be improved. In order to facilitate assessment, the researcher provided a model template so that students could complete the information with their evaluation in Google Drive without having to spend much time writing each message.

Additionally, the researcher evaluated every participant's podcast and offered feedback under the students' own posts. For the evaluation conducted by the researcher for each week's podcast, an evaluation sheet with exemplars of the target features from each week's podcast was created in a word document and printed out for every participant. The researcher marked aspects that were pronounced correctly in green and aspects that could be improved in blue (Figure 26). Then, a summary from this evaluation was posted on Edmodo, under each participant's podcast (Figure 27). This procedure provided students with feedback from three different people. Participants were supposed to look at this feedback and then try to improve those aspects in the next podcast.



6				today; are; about; the; has; familiar; are; temper; just; that; annoy; (and); at; that; our; modern; angrier; and; today; question; a; the; that; annoyed; a; centre; annoying; of; the; annoys; that; and; a; about;	We; the; you; me; my; that; me; at; you; me; it; was it; for; you; you; you; me; Rob; us	1 ✓ ✗ 2 ✓ ✗ 3 ✓ ✗ 4 ✓ ✗ 5 ✓ ✗ 6 ✓ ✗	
10				today; are; about; the; has; familiar; are; temper; just; that; annoy; (and); at; that; our; modern; angrier; and; today; question; a; the; that; annoyed; a; centre; annoying; of; the; annoys; that; and; a; about;	We; the; you; me; my; that; me; at; you; me; it; was it; for; you; you; you; me; Rob; us	1 ✓ ✗ 2 ✓ ✗ 3 ✓ ✗ 4 ✓ ✗ 5 ✓ ✗ 6 ✓ ✗	
12				today; are; about; the; has; familiar; are; temper; just; that; annoy; (and); at; that; our; modern; angrier; and; today; question; a; the; that; annoyed; a; centre; annoying; of; the; annoys; that; and; a; about;	We; the; you; me; my; that; me; at; you; me; it; was it; for; you; you; you; me; Rob; us	1 ✓ ✗ 2 ✓ ✗ 3 ✓ ✗ 4 ✓ ✗ 5 ✓ ✗ 6 ✓ ✗	BBC /B/

Figure 26. Weekly evaluation sheet for the podcast dealing with weak forms and contrastive stress

In an effort to motivate students, a common gamification strategy was adopted (see Zicherman & Cunningham, 2011). Students were awarded gold, silver or bronze trophies as they completed tasks. This is one of the functions offered by Edmodo, especially helpful in order to stimulate students to continue participating actively in the study, and therefore, obtaining more trophies. Trophies were customised to represent the target feature students were working on during that week (see Figure 28 for an example). To control participation, the researcher kept a record of students' completion of tasks, monitoring whether learners had uploaded their podcasts, completed their self-evaluation, and the two peers' evaluation.

Jonas Fouz · 27 de abril, 2014

Hi!

1. Reduction of full vowels to schwa in unstressed positions: Good in many cases but you need to improve this a little. You've pronounced many instances very well (e.g. about, temper, just, the, etc.). As we will see next week, when words are unstressed, vowels are commonly reduced (usually to schwa). However, in many instances you have pronounced full vowels (similar to the Spanish vowels), but they should have schwas (e.g. hAs, fAmiliAr, Annoy, modErn, thAt, fOr). This is very easy to improve, we'll continue working on this next week. I'm sure that with just a few minutes you'll get it perfect! 😀
2. Capacity to unstress words that don't need to be stressed: brilliant in most cases. There are only 2 examples that you have stressed and shouldn't be stressed ('you' in 'when you feel someone'...; 'you' in 'we'll find out if you're right...').
3. Use of contrastive intonation: Very good. You've understood the point and you can modify the nuclear accent very well to emphasise a contrast. However, you haven't stressed the correct word in three instances (the 2 sentences in 2; the first sentence in 3; the sentences in 6). This may be because you haven't realised what the intended contrast was, but the difficult part (which is shifting stress) is good. Think about that and I'll give you the answers with the next podcast.
4. Other aspects you want to comment on: Careful with the pronunciation of 'annoying', you use the sound in 'John' /dʒ/ for the 'y', but it's a diphthong + /ɪ/ ==> /əˈnɔɪ.ɪŋ/.

Also, about the pronunciation of 'push' and 'shove', you've pronounced <sh> with /z/ (which is a really difficult sound to pronounce, well done in that sense! we'll cover that in the last podcast), but the sound you need is /ʃ/.

Great job!

Figure 27. Sample of the feedback provided by the researcher under the student's post in Edmodo

Figure 28. Sample trophies and badges awarded to students for their participation

In order to illustrate the procedure visually, Figure 29 shows the calendar students were required to follow each week. Podcasts were sent on Monday and they could complete the first (input) stage until Thursday. The recording stage (output) had to be completed by Friday so that all participants had time to evaluate each other's podcasts before recording the next podcast the following Tuesday (evaluation stage).

Day	MON	TUES	WED	THUR	FRI	SAT	SUN
Task	Listen	Listen	Listen	Listen			
		Record	Record	Record	Record		
					Evaluate	Evaluate	Evaluate
	Evaluate						

Figure 29. Calendar students received with the three weekly stages they had to complete

4.1.3 Results and discussion

As noted above, this pilot study was conceived of as a scaffold for study 1. The data from the perception and production tests have not been analysed and therefore they will not be reported here. Nonetheless, since the results obtained from this pilot study guided the planning of the second study, some general findings considered significant for the implementation of a project of this kind will be discussed here. These not only guided the design of the second study, but also offer implications for teachers and researchers that wish to incorporate podcasts into their FL classrooms. Hence, the remainder of this section will briefly address three main points: the learners' engagement in the study, their responses to the questionnaires sent before and after training, as well as some issues that can be considered problematic for future implementations of similar projects in FL classrooms.

The students' involvement in this study was far from satisfactory. As explained above, students had to follow three different steps every week. Firstly, students had to read a PowerPoint presentation with explanations about the problematic aspect of that week and to listen to a podcast that provided examples of those features; secondly, they had to record their own podcast and upload it onto Edmodo so that other classmates would be able to access it; and finally, they were asked to self-evaluate their pronunciation of the target features in their podcast and provide peer-feedback to two other students. A register of task-completion shows that participants did not even follow the researcher's instructions in week one. Figure 30 shows the degree of completion of different tasks by participants in the experimental group. The graph shows that 26 students started recording the first podcast, but the number quickly decreased to 22 in the second week, 14 in the third week, 13 in the fourth, and finally 16 in the last one.

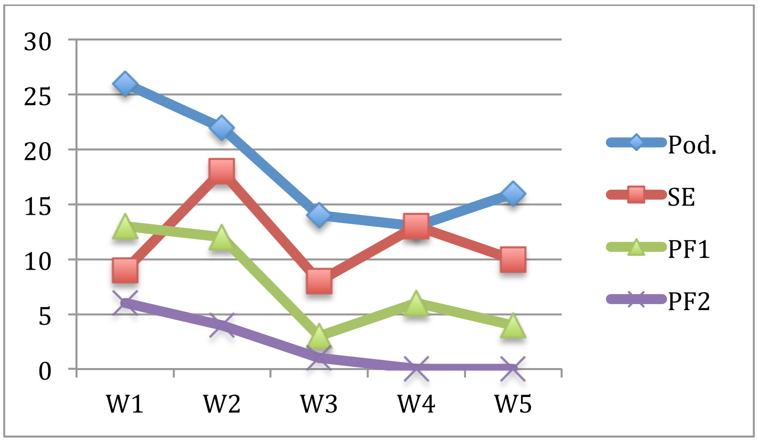


Figure 30. Number of participants that completed each task per week (W) – podcast (Pod.), self-evaluation (SE), and peer feedback (PF) 1 and 2.

While 26 students recorded the first podcast in week one, only nine completed their self-evaluation questionnaire, 13 completed the first peer feedback, and only six evaluated the second classmate. In the second week, 22 students recorded the podcast and 18 completed the self-evaluation questionnaire; nonetheless, only 12 evaluated the first peer, and only four evaluated the second peer. As the number of podcasts recorded every week decreased, the number of peers available to be evaluated was consequently smaller too.

This low level of engagement was very problematic for the training procedure, as learners were assigned their peer-assessors at the beginning of the week and they would not be able to evaluate or be evaluated by participants who were absent. This could explain the low number of peer evaluations received from week 3 on, where there are only three evaluations for peer 1 in the third week, six in week 4 and four in week 5. It could be the case that participants recorded their podcast and were waiting for their peers to upload theirs so that they could evaluate them but their assigned peer failed to do so. However, despite the fact that this may have been inconvenient in order to perform the peer evaluations, many students failed to complete their self-evaluations after recording their podcasts. This cannot be attributed to the peer assignment at the beginning of the week, as learners could perform their self-evaluations immediately after recording their podcast. The number of self-evaluations only coincides with that of the podcasts recorded in week 4. In all the other weeks, there are fewer self-evaluations than podcasts were recorded. This led the researcher to free students from sending the second evaluation and recommending them to evaluate the first person they saw without feedback on Edmodo. This is the reason why there are no evaluations for the second peer in weeks 4 and 5.

One possible explanation for the learners' low participation rate could be that they did not perceive the method or the contents to be useful. Nonetheless, their responses to the 5-point Likert-type questions in post-test questionnaires indicate that they considered the method (mean score 4.5; SD = 0.5)[83] and the contents (MS = 4.4; 1.1) useful for pronunciation instruction and applicable to their daily use of English (MS = 4.5; 0.7). Furthermore, their responses suggest that they considered that the method had helped them improve their pronunciation of the target aspects (MS =4.3; 0.7), made them more aware of their pronunciation and their mistakes (MS = 4.6; 0.7), and raised their awareness about the way people talk and pronounce English (MS = 4.5; 0.7).

Given the learners' positive responses regarding the perceived usefulness of the approach and the contents, the low participation rates may be due to learners finding the different weekly steps too demanding. Of the 15 learners in the experimental group who completed the final questionnaires, nine (60%) answered 'no' when asked whether the study was too demanding, two (13.3%) replied 'sometimes', and two others replied 'yes'. One student (6.6%) claimed that overall, (s)he did not find the tasks demanding, but that the second peer evaluation was too much. Finally, another participant said that tasks were OK, but claimed not to have been able to keep up with all of them due to other commitments.

The self and peer evaluations were considered to be an essential part of the process, as they drew the learners' attention to pronunciation and were therefore considered as an initial step for self- and peer-monitoring. Nevertheless, it may also be the case that participants did not perceive them to be as useful and therefore did not devote the necessary attention to them. Several questions in the final questionnaires dealt with the peer feedback learners gave and received (questionnaires can be found in appendices 4 and 5). This was done in an attempt to canvass their perceptions on their ability to provide feedback, evaluate their peers' pronunciation and their own, but also their perceived level of engagement when giving and receiving feedback. When asked about the adequacy of the feedback received by their classmates, they rated it with a mean score of 3.8 (1.3). However, they rated themselves more harshly when asked to reflect on the feedback they gave to their peers and whether they thought it had helped them. The mean score for the

[83] For the sake of simplicity, standard deviations (SD) will be presented in parentheses immediately after mean scores (MS).

perceived usefulness of the feedback they offered was 3.2 (1.2), which decreased to 2.9 (1) when rating their ability to assess their classmates.

As regards the amount of podcasts used and the length of the instruction, 14 students (93.3%) claimed that they found one podcast per week adequate and one said that it was too much. Nonetheless, all of them agreed that the length of the study (5 weeks) was appropriate. Surprisingly, this does not correlate with the empirical measures of participation, that is, with the degree of task completion. In the open-ended questions where participants could offer comments and suggestions, one said that the study had taken place at a very busy time of the academic year and that (s)he had not been able to complete some tasks due to many other assignments and exams at university. In fact, an item addressing learners' perception of their own involvement in the study reveals that seven students (46.6%) acknowledged not having made an effort, with six students (40%) claiming their participation was 'average' and that it could have been higher and one (6.6%) rating it as 'low'. Nevertheless, six students considered their participation to have been 'high', and two (13.3%) 'very high'.

The researcher was able to keep a record of the learners' participation by registering the 'tangible' tasks they completed (e.g. the podcasts they recorded, or the number of self and peer evaluations received). However, since training was conducted entirely online, it was not possible for the researcher to ensure that learners read the information sent in the PowerPoint slide show presentations (i.e. the explicit instruction) or the peer evaluations and the researcher's feedback they received. Thus, two of the items in the questionnaire asked learners whether they had read this information before recording their podcasts. Of the 15 students, eight (53.3%) claimed to 'always' read the information sent by the researcher, four (26.6%) claimed that they read it 'sometimes' and one (6.6%) acknowledged to 'never' have read it. Furthermore, while thirteen students (86.6%) claimed to 'always' read the researcher's and peer feedback before recording each week's podcast, one responded that (s)he only did it 'sometimes' and one that he 'hardly ever' did it.

The above-mentioned aspects were considered to severely affect the effectiveness of instruction. The explicit information sent at the beginning of each week was considered to be vital, as it was the element that was supposed to draw the learners' attention to form and foster noticing of the target features. Similarly, the weekly feedback offered by the researcher and by peers was supposed to help learners improve their pronunciation of the target aspects for the following podcast. Consequently, this strongly conditioned the final design for study 1. In order to make sure that all learners actually received instruction, some of the tasks in study 1 were conducted in class.

4.2 Study 1

In light of the findings obtained from the pilot experience described above, some modifications were made in the design of study 1. Even though the perception and production data from the pilot study have not been analysed, the learners' responses to the questionnaires and the limitations encountered served to pave the way for a better planned and more controlled study.

As discussed above, research in CAPT has often been conducted under very laboratory-like conditions that bare little resemblance to the way technology would be used in real implementations of the proposals under investigation. If CAPT is to enhance classroom instruction, conducting research in a classroom-like environment should not be problematic, as the environment would be similar to the one in which learners would actually use the technology. Nevertheless, for technologies that lend themselves to autonomous use, studies should assess their potential in the same conditions in which they would be implemented. The pilot study described above was an attempt to do this, but there were important limitations that seriously affected the evaluation of the approach.

The fact that learners did not read the theory that was sent to them when introducing each aspect or that they did not evaluate their peers' recordings does not permit a reliable evaluation of the approach. The theory offered for each aspect was one of the core elements of the instruction, consisting of the explicit information supposed to facilitate noticing. Similarly, the self- and peer-evaluation was supposed to raise the learners' awareness of others' speech and help them get used to monitoring pronunciation. By skipping these two steps, learners were simply not following training. Ideally, this type of training could be done from home and learners would be motivated enough so as to do it without external rewards (see e.g. Mompean & Fouz-González, in press), but given that students who participate in this type of study (receiving external rewards) may not be especially motivated per se, a more strict control was considered essential.

4.2.1 Aims, research questions and hypotheses

Study 1 focused on the two target aspects explained in section 2.4.1. These are: a) the distinction between English /s/ and /z/, and b) the pronunciation of /b d g/ as stops in positions where Spaniards would usually spirantise them. Thus, this study addresses three of the specific objectives in this dissertation, namely to investigate:

1. The potential of podcasts to improve Spanish learners' perception of the English /s – z/ contrast and of English voiced stops /b d g/ in intervocalic position.
2. The potential of podcasts to improve Spanish learners' production of English /z/ and to help them avoid spirantising /b d g/ in intervocalic position.
3. The learners' perceptions towards using podcasts for pronunciation training.

To address these three specific objectives, study 1 presents four research questions and seven hypotheses. These are based on the above-mentioned advantages of podcasts (sections 1.3.3 and 2.4.1) and the role of noticing and explicit instruction in SLA, and in particular, in pronunciation acquisition (section 1.1.5).

RQ1: Can learners' perception of the English /s – z/ contrast and /b d g/ as stops in intervocalic position be improved as a result of training with podcasts?

H1: Training with podcasts will help learners improve their perception the English /s – z/ contrast.

H2: Training with podcasts will help learners improve their perception of English /b d g/ as stops stops in intervocalic position.

RQ2: Do learners' improve their pronunciation of English /z/ and /b d g/ as stops in intervocalic position after training with podcasts?

H3: Training with podcasts will help learners improve their production of /z/.

H4: Training with podcasts will help learners avoid spirantising English /b d g/ in intervocalic position.

RQ3: Does training with podcasts generalise to novel words?

H5. Instruction with podcasts will help learners improve their perception of the /s – z/ contrast in novel words.

H6. Instruction with podcasts will help learners improve their production of /z/ in novel words.

H7. Instruction with podcasts will help learners avoid spirantising /b d g/ in intervocalic position in novel words.

RQ4: What are the students' reactions towards using podcasts for pronunciation training?

4.2.2 Method

The experience in the pilot study led to a considerable reduction of the duration of study 1 as well as of the amount of tasks learners had to complete weekly. The pilot study was designed with the intention of offering as wide a selection of aspects as possible. However, probably due to its duration, there was an undesirably high number of dropouts after week 3, which led to the decision of limiting instruction to three weeks in study 1. Since learners would receive a total of approximately 3 hours of training (one hour per week), it was considered that including too many different aspects would also be too demanding. As a case in point, González-Bueno (1997) devoted approximately the same time to help English learners of Spanish with the pronunciation of voiceless and voiced stops (/p t k/ and /b d g/), although training consisted of three 50-minute sessions every week in her study.

a) Participants

Forty-seven participants took part in this study, 35 female and 12 male (mean age 19.4; SD = 0.66). Given the poor participation rates obtained in the pilot study, possibly partly because of the proximity to the exam period, participants for studies 1 and 2 were recruited at the same time, at the beginning of the first (autumn) semester of the academic year 2014/2015. They were recruited from the same degree in English studies at the University of Murcia. Study 1 was advertised as part of the phonetics course students were enrolled on. Given that some students had already participated in the pilot study (as they were in the first year of their degree when the pilot study was conducted), an attempt was made to allocate students who had already participated in the pilot study to study 2. This was done to avoid them repeating the same type of instruction with podcasts. Nevertheless, in some cases, those learners who participated in the pilot study as control group were assigned again to study 1, this time as part of the experimental group. Assigning students to the same type of treatment could yield unreliable results regarding the effectiveness of podcasts, as some would have more experience with the procedure and may have advantage over others despite the fact that instruction was different this time. Nonetheless, given that subjects belonging to the control group in the pilot study did not receive any instruction with podcasts, they were allowed to participate in study 1.

As mentioned above, a possible explanation for the low participation rates in the pilot study might have been the fact that learners perceived they were going to receive the same score irrespective of the amount of work they had to do. Hence, participants in studies 1 and 2 were informed that they would all receive the same mark as all of them would receive instruction regardless of the group they were in. In this respect, as pointed out above, participants in study 1 were divided into two groups that received instruction through podcasts at the same time, but each group with a focus on two different aspects. This allowed using both groups as experimental and control groups.

One group received instruction on the English /s – z/ contrast (group 1; *n* = 25) and the other was trained on the voiced plosives /b d g/ as non-approximants in intervocalic position (group 2; *n* = 22). It must be noted that since these students were enrolled on a phonetics course and they had already taken an oral skills course the previous year, they should be aware of some of the aspects introduced in the theory for each group, or they might have covered them during the study. This implies that even though learners act as control group for one of the two issues under study, slight improvements may occur in their pronunciation of those aspects as a result of their current course.

Given the importance of phonetics in the participants' degree, and given the suitability of both studies for the course contents, studies 2 and 3 were included as an integral part of the phonetics course that year, awarding 0.7 points of the final mark to those students who wished to participate. It substituted one of the assignments students had to complete for the course. Those who chose to participate in the study did not have to submite that assignment and were evaluated according to completion of the tasks in the study instead.

b) Instruments

The instruments used in study 1 were the same that were used in the pilot study: 'BBC 6 minute English' podcasts, Edmodo, and Google Drive online questionnaires (see section 4.1.2b).

c) Training procedure

The training procedure was structurally similar to that in the pilot study, although with some modifications in order to overcome the obstacles encountered in that pilot experience. The three-stage weekly approach (input – output – evaluation) used in the pilot study was slightly modified for study 1. Two major obstacles encountered in the pilot study were that some learners did not read the theory offered by the researcher before listening to the podcasts and that a great number of participants did not complete the weekly self- and peer-evaluations.

Given the lack of participation and the impossibility of controlling learners in the pilot study, training in study 1 was conducted in one of the computer rooms at university. This ensured that learners received the instruction offered by the researcher, listened to the weekly podcasts and evaluated their peers' podcasts. Thus, for study 1 participants were told that they would have to meet the researcher for one hour a week over a period of three weeks (i.e. a total of three sessions). In order to reduce the time demands imposed by the tasks in the pilot study, participants were not required to perform the weekly self-evaluation in study 1 and they only had to evaluate one podcast per week. The specific stages to follow were:

1) A brief explicit explanation about the key issues (theory presentation, in-class).

2) Listening to the podcast and completing two activities online (input, in-class).

3) Peer evaluation of the recordings from that week (evaluation, in-class).

4) Recording the next podcast (output, at home).

The stages each group had to complete weekly were structurally identical. The first session included only stages 1 and 2 because it was the initial session. Participants were given a brief explanation about the procedure for the following three weeks, a 10-minute presentation on the key aspects under study (i.e. the explicit instruction), and they completed activities 1 and 2 afterwards. Stage 3 could not be completed in that first session given that they had just received instruction on the aspects addressed, and therefore, they had not recorded any podcasts yet. In the following sessions, during the first 10 minutes learners were briefly reminded of the key features of the sounds under study and they were offered a summary of their results in activities 1 and 2 from the previous session (Figure 31). If students asked questions or requested clarifications, sample stimuli were played again and the researcher offered further explanations when necessary. Then, they listened to the next podcast and completed the two online activities on stage 2. Finally, they evaluated the podcasts their peers had recorded during the previous week.

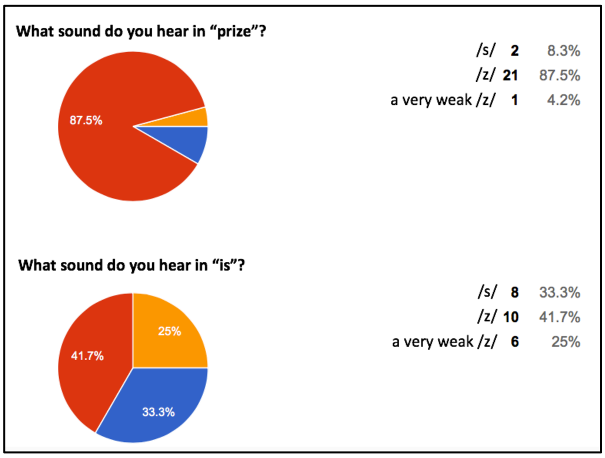


Figure 31. Summary of students' responses to activity 2 (input stage)

A new addition with respect to the pilot study was the inclusion of two activities on stage 2. This was done to make sure that participants actually paid attention to the target aspects when listening to the podcast. In the pilot study, participants were told to simply listen to the podcast and pay attention to the way these sounds were pronounced, but this cannot guarantee that they actually did so, or even that they paid attention to the right words. Hence, two activities were created in order to direct participants' attention to the key features. In activity 1, learners had to listen to a 3-minute version[84] of the podcast and write down at least 10 words for each of the target sounds (Figure 32). Activity 2 was aimed at assessing the learners' perception of the target sounds in specific words by means of multiple-choice questions comparable to the forced-choice tasks employed in perceptual identification tasks (Figure 32). Examples included instantiations of the same sound as represented by different spellings and in different positions (e.g. /s/ in *associated, just,* as part of /ks/ in *exposure*, or /z/ in *cause, opposite,* as part of /gz/ in *exactly*).

[84] The full 6-minute podcast includes a summary of key words in the podcast at the end, as well as a brief introduction by the BBC at the beginning with information about where to find more podcasts, etc. Podcasts were edited and reduced to three minutes in order to reduce the amount of time required for stage 1. This allowed participants to complete activity 1 by only listening to a reduced version of the podcast that contained a good number of exemplars. This has also been recommended by McBride (2009a), who points out that it is much easier to work with edited versions rather than full-length podcasts, which may range from eight minutes to an hour.

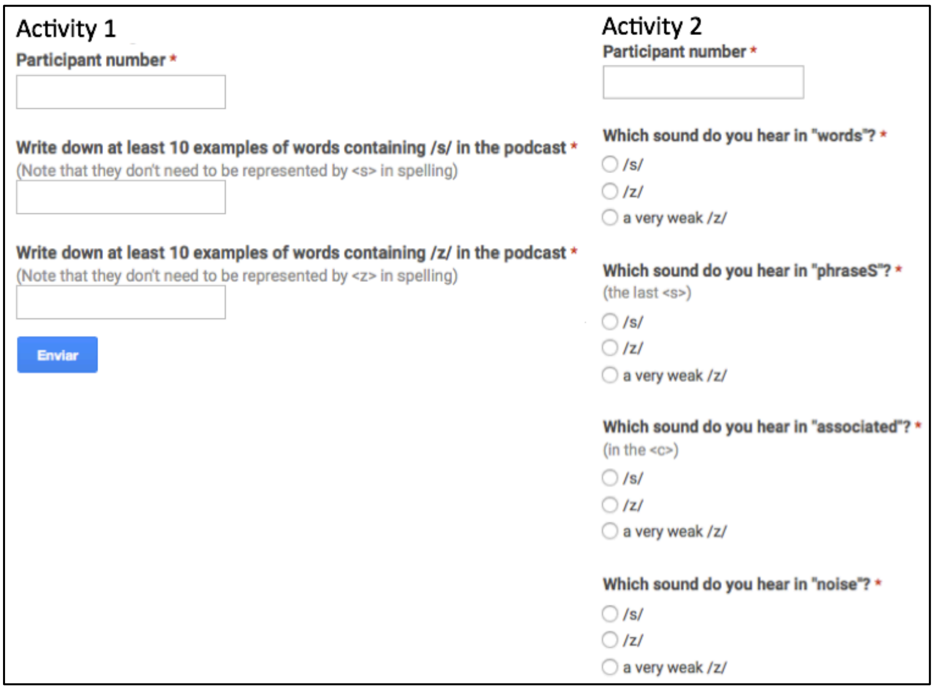


Figure 32. Sample activities for stage 2 for the group receiving instruction in the /s – z/contrast

As in the pilot study, podcasts for each week were uploaded onto Edmodo so that learners could access them in an individual computer in class and listen to the activities at their own pace. Even though this approach is much more controlled than that of the pilot study, it resembles the way in which this tool could be used from home. The researcher created two groups in Edmodo, one for each group in the study. This was done so that learners in one group did not have access to the materials for learners in the other group. Activities 1 and 2 were created using Google Drive's questionnaires, which allowed for an automatic collection of data on students' performance and participation. For activity 2, podcasts were divided into shorter excerpts that were uploaded onto Edmodo as separate files (Figure 33). Each excerpt had a link to a different questionnaire, which allowed a fast and convenient way of focusing on the target sounds without requiring participants to spot the target word in the whole podcast.

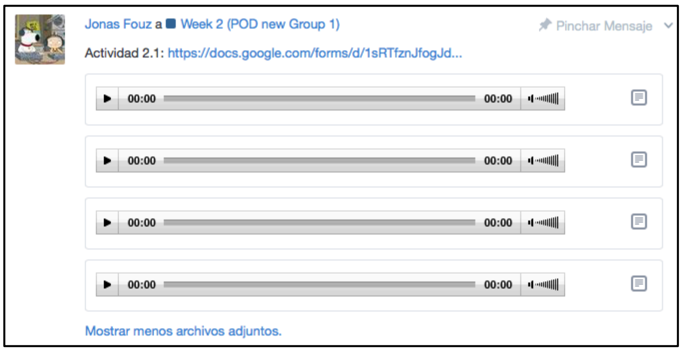


Figure 33. Sample image of activity 2 (input stage) as seen in Edmodo for group 1

The next stage (stage 3) was completed after the activities on stage 2. However, as noted above, this was only done on weeks 1 and 2 of the study given that learners had not recorded any podcasts during the first week. No text was sent after the third week as learners were not required to meet with the researcher for a fourth session and this would change the procedure followed in the previous two sessions (since learners would have to complete their evaluations from home). The procedure for the peer evaluation was the same as for activity 2 (Figure 32, right). Learners had to complete a questionnaire focusing on specific examples selected by the researcher. The evaluations were anonymous, but students were required to specify their peer's participant number so that they could check the evaluations given by their peers afterwards. The spreadsheet with all the participants' scores was uploaded onto Edmodo after the end of each week's training session.

The participants' responses to the questionnaires in the pilot study revealed that the explicit feedback offered by learners was not perceived to be useful in that learners were not able to detect the causes of errors and provide explanations. Therefore, for study 1, participants were required to rate the learners' pronunciation following the same approach in activity 2. In other words, the rating was carried out by means of multiple-choice questions simply assessing whether the target had been met or not. Similarly, since the aim of study 1 was to explore the effect of podcasts for autonomous learning, although peer evaluations were offered, the teacher did not mark or provided feedback on their recordings.

Given the problems encountered in the pilot study, instead of assigning participants for peer review, participants in this study were encouraged to assess the first recording they saw without evaluation. Since they were all engaged in the evaluation at the same time, they were instructed to post a comment under their peer's podcast stating that they were evaluating that recording. This was a way of marking that the podcast at hand was already being evaluated and avoiding an uneven number of evaluations at the end.

As regards stage 4, the students' own recordings, a text with a substantial number of instantiations of the target sounds (described below) was uploaded onto Edmodo so that students could rehearse and record their podcast from home. The texts were typographically enhanced, by underlining the orthographic representations of the target sounds and marking them in bold type. This type of typographical enhancement can be used in order to make input more salient (Han et al., 2008). However, in this case it was used as a means of output enhancement; that is, to draw the learners' attention to a particular sound in order to facilitate their production. Nonetheless, learners were told that they would have to find out about the pronunciation of each particular example themselves. The intention was to draw their attention to the different possible orthographic representations for each sound.

d) Training stimuli

This section describes the stimuli and materials employed for the four weekly tasks used for groups 1 and 2. The materials for each group will be described in the order followed for the completion of the tasks: (1) theory presentation; (2) input provision (activities 1 & 2); (3) peer evaluation; and (4) recording.

Stage 1: Theory presentation.

The explicit explanations on the English /s – z/ contrast dealt with the place and manner of articulation of these two sounds, including diagrams that exemplified the articulatory movements required for both sounds. Learners were explicitly informed about the common devoicing of /z/ in certain contexts and the lengthening effect it has on the preceding vowel. However, they were encouraged to add voicing in order to mark the distinction between /s/ and /z/. The most common spellings for both sounds were also presented, including a brief explanation on the contexts in which they are likely to occur, accompanied with examples (e.g. the –s suffix in plurals, third person singular, Saxon Genitive and contractions pronounced with /s/ after voiceless sounds – as in *works, writes, maths*; and /z/ after voiced sounds – as in *flies, buys, needs, cars, clothes*). These two sounds were also compared with the voiceless post-alveolar fricative /ʃ/, which was included as distractor in some examples of activity 2.

The explicit explanations for English /b d g/ dealt with the different stages in the articulation of stops, accompanied with diagrams and animations. Explanations included information about the different degrees of constriction for English consonants (stops, fricatives, approximants, etc.), and differences in place and manner of articulation (bilabial, alveolar and velar) for English stops. Participants were explicitly informed about the differences in manner of articulation for English /b d g/, as compared to their Spanish counterparts, which are typically spirantised in intervocalic position. They were also informed about the differences in place of articulation for the English alveolar plosives /t d/, which are dental in Spanish.

Stage 2: input provision.

As noted above, podcasts were selected from the 6-minute English series by the BBC. The three podcasts included for training were selected after an examination of a range of podcasts by the same series. The criteria for the inclusion of a podcast as part of the training materials were that it included a considerable number of instantiations of each target sound and that exemplars of target sounds in the podcast could be considered to be 'prototypical' (i.e. good examples of the sound – see Mompean, 2001, 2004a).

The 6-minute English series offers free scripts for every podcast. The scripts were automatically transcribed with the PhoTransEdit software (PhoTransEdit, 2013) in order to calculate the number of occurrences of the target sounds in the podcast. This programme enables users to transcribe text automatically, offering a summary with the number of examples of each phoneme in the text (Figure 34). A list including all the potential examples for each sound (based on their spelling) was compiled for every podcast. An effort was made to include a balanced number of examples of the target sounds. The scripts for the podcasts used with both groups can be found in Appendix 6.

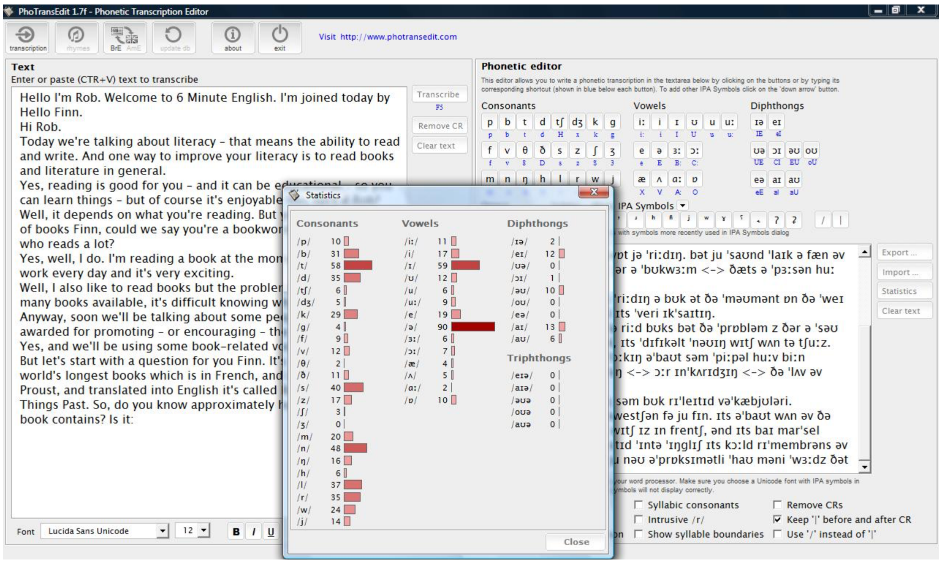


Figure 34. Screenshot of automatic transcription and summary of instantiations of the sounds in a sample excerpt in PhoTransEdit

It must be noted that although an effort was made to find podcasts featuring the target sounds in different contextual positions, training materials were subject to the availability of existing podcasts. In this respect, unlike experimental studies in which stimuli are created by the researchers, studies implementing already-existing materials (as is the case of studies 1 and 2) do not always allow researchers/teachers to control for the different contexts in which target sounds and patterns occur.

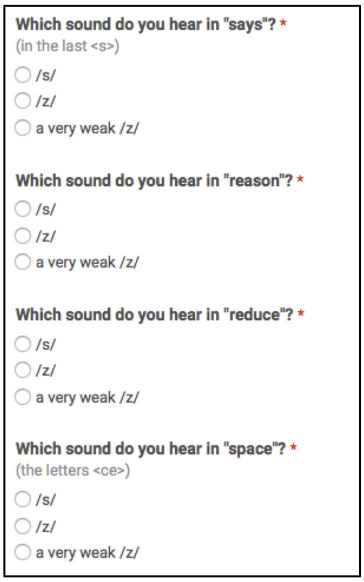
Hence, the target sounds in these podcasts appeared in a number of phonetic contexts that represent the natural variability occurring in authentic input. Contexts for /s/ featured the sound both in singleton position and in consonant clusters with the sound spelt in different ways. The sound occurred in word initial (e.g. *survey, city, start, smaller, skilled, space*), medial (e.g. *associated, exposed, buses, possibly*) and final position (*peace, course, groups, lengths, less, lots*). Its voiced counterpart (/z/) also appeared in different phonetic environments and spelt with different graphemes, although limited to word-medial (*phrases, noisy, example, closed, wasn't, prizes*) and word-final position (*noise, words, jobs, buses, quiz, drills*).

Table 1 shows the number of exemplars of each sound per podcast, including the percentage of appearance in relation to the total number of target sounds in every podcast (i.e. the frequency of occurrence of one sound in relation to the other), as well as the total number of examples across podcasts. The whole list of words containing the target sounds in each podcast for groups 1 and 2 can be found in Appendix 7. As shown in Table 1, the rate of occurrence of the two target sounds in each podcast was rather similar, with a maximum difference of 13 points between the two sounds in podcast 3.

Table 1. Number of examples and percentages of appearance of each sound for /s – z/per podcast (Pod.) and in total (Tot.)

	Pod. 1		**Pod. 2**		**Pod. 3**		**Tot. per sound**	
	nº	**%**	**nº**	**%**	**nº**	**%**	nº	**%**
/s/	74	47.4	81	54	106	56.7	263	53.1
/z/	82	52.6	69	46	81	43.3	232	46.9
Tot.	156		150		187		493	

As noted above, activity 1 on this stage required learners to find around 10 examples of each sound in the podcast and write them down. Activity 2 was similar to an identification task in which they had to decide which of the two sounds they were hearing in a selection of examples. Stimuli included exemplars of both sounds spelt with different graphemes, occasionally including two questions for the same word where both /s/ and /z/ appeared. This was done to raise learners' awareness of the different possible spellings for these sounds (Figure 35). Activity 2 for the /s – z/ contrast included a third 'weak /z/' option because there were several examples of /z/ in final position in which the sound was partially devoiced but, nevertheless, an instance of /z/.[85]


Which sound do you hear in "says"? *
(in the last <s>)
/s/
/z/
a very weak /z/

Which sound do you hear in "reason"? *
/s/
/z/
a very weak /z/

Which sound do you hear in "reduce"? *
/s/
/z/
a very weak /z/

Which sound do you hear in "space"? *
(the letters <ce>)
/s/
/z/
a very weak /z/

Figure 35. Example of identification task in activity 2 for group 1

Regarding /b d g/, only those examples where the sounds were preceded by vowels were counted as instances of the target aspects. Table 2 shows the number of examples of each sound per podcast, including the rate of occurrence in relation to the total number of

---

[85] Despite the fact that /z/ is typically partially devoiced in word-final position, other phonetic cues distinguish it from /s/, such as preceding vowel length (Shockey, 2003) or its their comparatively shorter and lesser degree of frication (Brown, 1990).

target items in each podcast and the total across podcasts. The whole list of lexical items showing the target sounds for each podcast can be found in Appendix 7.

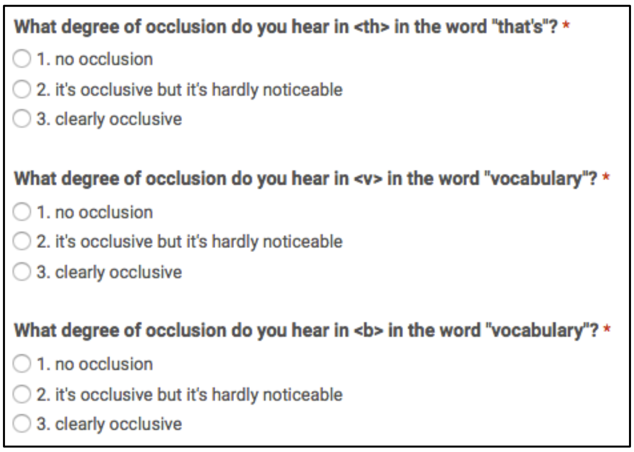
Table 2. Number of examples and percentages of appearance of each sound for /b d g/, per podcast (Pod.) and in total (Tot.)

	Pod. 1		Pod. 2		Pod. 3		Tot. per sound	
	nº	%	nº	%	nº	%	nº	%
/b/	50	37.9	49	29.5	70	32.3	169	32.9
/d/	61	46.2	73	44	115	53.5	250	48.6
/g/	21	15.9	44	26.5	31	14.3	96	18.7
Tot.	132		166		216		514	

As with the materials for group 1, an attempt was made to include podcasts that were balanced in terms of the number of occurrences of each of the target sounds. In this regard, of the three target sounds, /d/ was the most frequent in the three podcasts. In comparison with the two other sounds, /g/ was relatively underrepresented. Nonetheless, it was believed that since the nature of the target aspect for group 2 is not a problem of sound transfer, but of the Spanish spirantisation rule in intervocalic position, all three sounds serve to illustrate the target pronunciation, as they all need to be articulated as stops when flanked by vowels. Similarly, orthography should not pose problems for the pronunciation of these sounds, as the phonetic symbols for those sounds are the same as their graphemes. Thus, despite the fact that there were more instantiations of /b/ and /d/ than of /g/ in the podcasts, this was not considered a problem given that the three sounds help to illustrate the target pronunciation.

As with the target sounds for group 1, items for group 2 also occurred in various phonetic contexts (see Appendix 6). These included initial, medial, and final position (preceded and followed by vowels in initial and final position). Tokens featured the sounds both in singleton position and as part of consonant clusters (*British, blame, drink, words, group, glass*).

Activities 1 and 2 were identical in format to those for group 1. Figure 36 shows three sample questions from activity 2. Learners were asked about the degree of occlusion they perceived in a selection of examples. Examples included instantiations of the target sounds as taps (e.g. for /t d/ in unstressed positions), but also other sounds that may be confused with English /b d g/, such as the voiced labio-dental fricative /v/ (often confused with /b/), or the voiced dental fricative /ð/ (often confused with /d/) – see Monroy (2001). Introducing comparisons with these sounds was considered to be a good way of helping participants notice the differences between fricatives and stops, which is what they need to become aware of in order to avoid spirantising /b d g/ in English.


What degree of occlusion do you hear in <th> in the word "that's"? *
- 1. no occlusion
- 2. it's occlusive but it's hardly noticeable
- 3. clearly occlusive

What degree of occlusion do you hear in <v> in the word "vocabulary"? *
- 1. no occlusion
- 2. it's occlusive but it's hardly noticeable
- 3. clearly occlusive

What degree of occlusion do you hear in <b> in the word "vocabulary"? *
- 1. no occlusion
- 2. it's occlusive but it's hardly noticeable
- 3. clearly occlusive

Figure 36. Sample activity for group 2 (activity 2)

Stage 3: Peer evaluation.

As noted above, the peer evaluations were also carried out using Google Drive, following the same format as in activity 2. The researcher selected a range of words featuring the target sounds from the texts students had to record and created the questionnaires. This facilitated learners' focusing on those aspects, but at the same time it allowed all learners to be evaluated according to their pronunciation of the same target words. At the end of each week's session, the summary of the peer evaluation was shared through Edmodo so that learners could see their scores for each of the target words and work on the aspects they needed to improve before the following week's training.

Stage 4: Recording podcasts.

In the pilot study, the texts participants had to record were excerpts from same podcast they received as input. Nevertheless, since this imposed considerable time demands, it was decided to make this stage shorter. In study 1, two texts were created for learners to record at home during weeks 1 and 2. As explained above, no text was included for week 3 given that the study only lasted 3 weeks. Texts were designed with the aim of including as many examples as possible of the target sounds in order to foster production practice with the target features. The texts and the target words that served as prompts for groups 1 and 2 can be found in Appendix 8. This includes the number of instantiations of each target sound in each text and their percentage of occurrence with respect to one another.

e) Testing procedure

As noted above, the two groups in this study acted as both experimental and control groups. Group 1 received training on the English /s – z/ contrast and group 2 on the realisation of English /b d g/ as non-approximants in intervocalic position. Despite the fact that they received different types of training, both groups took the same pre- and post-tests. This allowed measuring potential improvements on the items on which each group was trained, but also to control whether improvements were made for the group that did not receive training.

Pre- and post-tests consisted of a total of four meetings with the researcher. The production tests were administered individually with the researcher and lasted around ten minutes (10 mins x 47 students x 2 testing times = 15 hours and 40 minutes approximately). The perceptual tests were administered in groups of around 17 students. There were four shifts both in each testing time (pre and post), two in the morning and two in the evening so that learners could choose a time of their convenience. The duration of the perceptual tests was approximately 40 mins (40 mins x 8 shifts = 5.3 hours).

Perception of the /s – z/ contrast was measured with the two perceptual tasks described in section 3.1.2. The perception tests were divided into two main tasks: a discrimination task and an identification task. However, these tasks were further divided into several activities in order to test the different target features. In this study, there was a discrimination task for the /s – z/ contrast and two identification tasks, one for the /s – z/ contrast and another one for /b d g/. Perception of /b d g/ was also evaluated through a delayed mimicry task, although this was administered as part of the production test (Figure 37 illustrates the arrangement visually). The production test consisted of three tasks, an imitation task, a sentence-reading task and a timed picture-description task (a description of these tasks is offered in sections 3.1.2 and 3.1.3).

The discrimination test was created in TP (Rato et al., 2015) as four separate tasks. The first was a training session in order to familiarise participants with the aural oddity task. It consisted of 10 triads with minimally-paired words featuring non-target phonetically distant sound contrasts (e.g. /p – l/ *pimm-limb-pimm;* /p – s/ *poon-soon-soon;* /p – tʃ/ *pin-pin-chin*), including catch triads (e.g. *talk-talk-talk*). The second task addressed the target contrast in initial position (e.g. *zap-sap-zap*), the third in medial position (e.g. *fuzzy-fussy-fussy*), and the last one in final position (e.g. *prize-price-price*).

(1) Discrimination task: /s – z/ contrast

a. Activity 1: training session with non-target stimuli
b. Activity 2: discrimination of /s – z/ in initial position
c. Activity 3: discrimination of /s – z/ in medial position
d. Activity 4: discrimination of /s – z/ in final position

(2) Identification task: /s – z/ contrast and occlusion of /b d g/

a. Activity 1: identification of /s/ and /z/
b. Activity 2: identification of English and Spanish /b d g/ – same or different?

(3) Accent mimicry task: occlusion of /b d g/ (administered as part of the production test)

Figure 37. Structure of the perception pre- and post-tests for study 1

Before starting the test, just after the oddity discrimination training with geometric shapes (see section 3.2), learners were explained the different parts of the task and provided with some examples (e.g. for words with the contrast in initial position *lip-sip-lip;* for medial *riding-rhyming-riding;* for final *bus, but, but*). Emphasis was made on the need to read the instructions before starting every task, as each task required participants to focus on the contrast in a different position. No feedback was offered during the test, except for a summary of the number of hits and errors at the end.

With regard to the identification task, it was divided into two parts (Figure 37 above). The first one was aimed at measuring the participants' perception of the /s – z/ contrast and the second focused on /b d g/. After the discrimination task was over, there was a brief pause in which the researcher handed out a set of photocopies for the identification task for /b d g/ and gave instructions on how to proceed.

For the /s – z/ contrast, the typical format of a conventional identification task was used. Words were presented individually and learners had to decide which sound they heard among three options, including the target sounds (/s/ - /z/) and the distractor (/ʃ/). As in the discrimination task, an 'I don't know' option was included in order to avoid random responses. However, learners were told to use this option only in cases where they really could not opt for one of the options.

In order to measure the perception of English /b d g/, a task in which learners had to 'identify' sounds as /b/, /d/, or /g/ was not considered appropriate since the problem Spanish learners have does not stem from transfer of an L1 sound to the FL, but rather, from a transfer of the /b d g/ spirantisation rule in Spanish (Zampini, 1996). As explained in section 3.1.2, despite subtle phonetic differences, in theory, learners could identify examples of English and Spanish sounds as 'the same' phoneme and they would not be entirely wrong. Thus, learners were asked to decide whether the degree of occlusion produced in the articulation of a number of sounds was the same in Spanish and English.

Participants were given a list of Spanish words with the orthography corresponding to the target sounds underlined and in bold (see Appendix 9). Recordings of the English versions of those words were then played through headphones. Learners listened to one word at a time and, by reading the word in Spanish and pronouncing it to themselves, they had to decide whether the degree of occlusion in the pronunciation of the target sounds was the same in English and Spanish. The task was set up in TP as an identification task with the labels 'yes', 'no', and 'I don't know'. For example, learners heard the word *dagger* in English (not provided in its written form), and read the word *daga* in Spanish, having to decide whether the /g/ had the same degree of occlusion in both languages.

Finally, learners' were asked to imitate the way English speakers pronounce Spanish in a delayed accent mimicry task (see section 3.1.2). Although this task was aimed at measuring perception, it was part of the individual interviews for the production test, just after the picture-description task. The task was meant to assess participants' awareness of the realisation of /b d g/ as stops in intervocalic positions.

f) Testing stimuli

Perception

A list of 15 minimally paired words was compiled for the /s – z/ discrimination task. Ten pairs were aimed at measuring the /s – z/ contrast and five were included as distractors contrasting /s/ with the voiceless post-alveolar fricative /ʃ/ (see Table 3). Test items featured the target contrast in word-initial (n = 4), word-medially (n = 4) and word-finally (n = 2).

A total of 25 triads were created for the test, 10 were change triads and 10 were catch triads featuring the contrast in the above-mentioned distributions. The other five triads were used as distractors. The two target sounds appeared as the 'odd one out' the same number of times. Tokens in bold in Table 3 show the odd items out in the triads. Additionally, the position of the odd item was randomised so that there was a balanced number of correct responses under each label (1, 2, or 3) with the exception of catch triads whose response always had to be 'the same'. An effort was made so that every triad contained randomised male and female voices.

Table 3. Stimuli for the /s – z/ oddity discrimination task in study 1.

Change triads /s - z/			Catch triads /s - z/			Distractors /s - ʃ/		
1	**sap**	zap	**11**	sing	sing	**21**	see	**she**
2	seal	**zeal**	**12**	zap	zap	**22**	**sue**	shoe
3	said	**z**	**13**	zing	zing	**23**	seesaw	**seashore**
4	sing	**zing**	**14**	sap	sap	**24**	**Iris**	Irish
5	fussy	**fuzzy**	**15**	precedent	precedent	**25**	mass	**mash**
6	president	**precedent**	**16**	muscle	muscle			
7	**muscle**	muzzle	**17**	fuzzy	fuzzy			
8	**racing**	raisin	**18**	raisin	raisin			
9	**bus**	buzz	**19**	buzz	buzz			
10	price	**prize**	**20**	price	price			

As for the identification test, familiar words were selected from the most frequent words in the podcasts, although there were some instances in which words with a low number of occurrences were chosen because of the spelling they exemplified (e.g. /z/ as represented by <s>) or because of their position within the word (initial, medial or final). These words were also included in activity 2 during training and in the texts participants had to record weekly. As mentioned above, 120 words featured the sound /s/, and 92 /z/ in the three podcasts employed for training. Of those 212 words, the majority (81 for /s/; 57 for /z/) appeared only once throughout the podcasts. There were several words that appeared two or three times, and a few that appeared more than ten times (see Appendix 7).

Table 4 shows the ten words included as 'familiar' and 'novel' in the identification test and their frequency of occurrence in training. They are arranged according to their position in the word: initial (/s/ n = 3; /z/ n = 0), medial (/s/ n = 4; /z/ n = 5), and final (/s/ n = 3; /z/ n = 5). Eight words were included as distractors featuring the voiceless post-alveolar fricative /ʃ/. Although items are arranged in Table 4 according to their position in the word, the order of items featuring /s/, /z/ and /ʃ/ was randomised in the tests.

Table 4. Familiar and novel words included in the pre- and post-tests for /s – z/ including the number of occurrences of familiar words during training (NO)

	Familiar						Novel				Distractors
	/s/	NO		/z/	NO		/s/		/z/		/ʃ/
1	so	14	**11**	music	4	**21**	sign	**31**	zero	**41**	shoe
2	some	7	**12**	reason	4	**22**	soup	**32**	Zoe	**42**	she
3	sound	5	**13**	example	2	**23**	centre	**33**	Xerox	**43**	mission
4	person	3	**14**	exactly	1	**24**	ceiling	**34**	bosom	**44**	issue
5	answer	1	**15**	easy	1	**25**	circle	**35**	resume	**45**	bush
6	question	7	**16**	is	25	**26**	professor	**36**	treason	**46**	sugar
7	just	10	**17**	noise	24	**27**	December	**37**	desire	**47**	mansion
8	it's	12	**18**	as	9	**28**	decent	**38**	arise	**48**	passion
9	yes	5	**19**	has	7	**29**	mouse	**39**	erase		
10	office	6	**20**	was	6	**30**	close (adj.)	**40**	muse		

With regard to the identification task measuring the learners' perception of occlusion for /b d g/, the researcher compiled a list of 48 English and Spanish cognates containing examples of voiced and voiceless stops in different positions. An effort was made to include cognates featuring the target sound in stressed and unstressed syllables in both languages. There were some instances in which the sound was stressed in one of the languages but unstressed in the other (e.g. *labor* vs. *labour*). Nonetheless, stress was not considered to have an effect on the learners' perception of the sounds, L1 Spanish speakers tend to spirantise these sounds regardless of their stress pattern (see examples in Monroy, 2001). Even though the phonological structure of these cognates was not exactly identical (e.g. having a different vowel sound afterwards in some instances – *abadía* [aβa̱ˈðia] vs. *abbey* [ˈæbi̱]), the criterion was to include pairs of words perceived as being very similar, featuring the target sounds in intervocalic position.

Each target sound was featured in 10 different words, five in word-initial position followed by a vowel and five in word-medial intervocalic position. Thus, a total of 30 words were devoted to assessing the learners' perception of /b d g/ in word-initial and word-medial intervocalic position. Nevertheless, in order to avoid learner biases as a result of one position of the sound in the word (i.e. initial vs. medial), nine more items (three per target sound) were included as 'control items'. These featured the target sounds preceded by nasal sounds – contexts where, despite occurring word-medially, /b d g/ are pronounced as stops. These control items should be perceived as having the same occlusion as their English counterparts. Hence, words featuring the target sounds in initial position and preceded by nasals served as a way of testing whether participants actually based their responses on the degree of occlusion, rather than automatically assuming that the English and Spanish versions were different. Finally, other nine items were included as distractors and were not considered for the results of this test. These featured the non-target English voiceless plosives /p t k/ (see Table 5). Therefore, the total number of items in the test was 48. It should be noted that even though each target sound is represented in 13 words, all the 39 items serve to assess the learners' perception of /b d g/ as stops in intervocalic position, as the aspect under study is the same for the three target sounds (i.e. avoiding to spirantise stops in intervocalic position).

Table 5. Stimuli for the identification task assessing perception of English /b d g/ in study 1

	/b/		/d/		/g/	
	Spanish	**English**	**Spanish**	**English**	**Spanish**	**English**
Initial position	banana	banana	danza	dance	gol	goal
	banda	band	dólar	dollar	guía	guide
	bar	bar	daga	dagger	guitarra	guitar
	base	base	debate	debate	gas	gas
	bate	bat	defensa	defence	gorila	gorilla
Medial position	Ébola	Ebola	adicto	addict (v.)	daga	dagger
	abadía	abbey	crédito	credit	mega	mega
	labor	labour	idioma	idiom	rigor	rigour
	obeso	obese	líder	leader	legal	legal
	Iberia	Iberia	medio	medium	figura	figure
Control items (/b d g/ preceded by nasal)	símbolo	symbol	industria	industrial	distingue	distinguish
	combinar	combine	referéndum	referendum	lengua	language
	miembro	member	apéndice	appendix	singular	singular
	/p/		/t/		/k/	
Distractors (non-target items /p t k/)	**Spanish**	**English**	**Spanish**	**English**	**Spanish**	**English**
	pin	pin	tipo	type	clan	clan
	aparte	apart	té	tea	eco	echo
	iPod	iPod	tomate	tomato		
	pan	pan				

Finally, regarding the accent mimicry task, each target sound was featured in five Spanish words. The target sounds occurred either in word-medial intervocalic position or in word-initial position flanked by vowels (see Table 6). Words featuring the target items were embedded in five sentences. However, three more sentences were included as distractors featuring instances of English voiceless stops /p t k/, sounds that also tend to be mispronounced by Spanish learners of English (Flege & Hammond, 1982). The 15 tokens from each participant (5 items x 3 target sounds x 2 testing times = 30 items x 47 participants = 1410 tokens) were edited together with the results from the production task and were analysed by the three expert judges described in section 3.3.

Table 6. Words and sentences for the /b d g/ delayed accent-mimicry task

	Words					
	/b/	**/d/**	**/g/**	**/p/**	**/t/**	**/k/**
1	debate	idioma	hago	comprar	tienes	qué
2	global	Canadá	figura	Palencia	irte	comprar
3	autobús	medio	Lugo	primos	tomate	cuándo
4	abuelo	puedo	agua	Pamplona	tía	quieres
5	bebe	cuidar	gusta	pedirme	falta	que
	Sentences					
1	No me gusta el debate sobre el idioma en Canadá.					
2	Nos encontramos en medio del calentamiento global.					
3	Mi primo es conductor de autobús y vive en Lugo.					
4	Hago lo que puedo para cuidar mi figura.					
5	Mi abuelo bebe agua a todas horas.					
6	¿A qué hora tienes que irte a comprar tomate?					
7	Su tía de Palencia tiene tres primos en Pamplona.					
8	¿Cuándo quieres pedirme el dinero que te falta?					

Production

Following the same criteria used for the selection of training stimuli in terms of the frequency of occurrence in the podcasts and appearance of the sound in different positions, a list of 10 familiar items was compiled for the sentence-reading task in order to measure the learners' pronunciation of /z/. Tokens featured /z/ in medial (n = 5) and final (n = 5) position, the same contexts in which the sound appeared during training. As in the perceptual tasks, five words featuring /s/ were included as distractors. Additionally, 10 extra tokens featuring /z/ were included as novel words to test possible generalisation gains. Of the 10 novel words, six featured the sound in a novel context (word initial position). Stimuli for the sentence-reading task can be found in Table 7. The whole set of stimuli including the carrier sentences employed can be found in Appendix 10.

Table 7. Stimuli for the sentence-reading task measuring controlled production of English /z/

	/z/							/ʃ/
	Familiar	**NO**	**NO2**	**NOTX**		**Novel**		**Distractors**
1	music	4	1	1	**11**	zoom	**21**	shoes
2	reason	4	0	1	**12**	zoo	**22**	shower
3	example[86]	2	1	2	**13**	zipper	**23**	sugar
4	residential	1	1	1	**14**	zone	**24**	ship
5	easy	1	1	1	**15**	zebra	**25**	sheep
6	noise	24	7	1	**16**	zombie		
7	prizes	6	1	2	**17**	cousin		
8	offices	11	1	0	**18**	busy		
9	dollars	4	2	3	**19**	prison		
10	words	2	0	0	**20**	museum		

Note: The table also shows the number of occurrences of the word in the podcasts (NO), in activity 2 (NO2) and in the texts learners recorded (NOTX).

The same procedure was followed for /b d g/. Based on the number of occurrences of the target words during training and on the distribution of /b d g/ within those tokens, 45 items were selected for the sentence-reading task. Each sound was featured in 10 familiar words in intervocalic position (five word-medial and five in word-initial position preceded by vowels), and five novel words (also in intervocalic position, this time only word-medially).

Research suggests that spirantisation is stronger in items preceded by stressed syllables (Eddington, 2011; Ortega-Llebaria, 2004), although L1 Spanish speakers tend to produce some degree of spirantisation regardless of stress (see examples in Monroy, 2001). However, in order to offer a more comprehensive measure of this aspect, the 15 tokens per sound included target sounds in accented (n = 9) and unaccented positions (n = 6). For the stimuli in the sentence-reading task, see Table 8.

The words in which the sound appears in initial position were always flanked by vowels (e.g. *a bit; the benefits; I didn't; very difficult* – see Appendix 10 for the whole set of stimuli, including carrier sentences). The only exception in which the target sound is not entirely surrounded by vowels is *programme*, but it was included given that there were no more words in training (i.e. familiar) featuring the sound in medial position. However, this context (/gr/) also lends itself to the spirantisation of stop consonants. In fact, the word

[86] Even though the <x> in the word *example* can be pronounced as /ks/ or /gz/ (Jones et al., 2011), it was considered an interesting token in order to illustrate that <x> can be a potential spelling for /z/, as evinced in the intantiations of this token in the podcasts. Since the version with the voiced segment is the one learners were exposed to during training, the participants' pronunciation of this word was evaluated on whether they had assimilated this information or not.

programme was one of the examples of consonant substitution (/g/➔ [ɣ]) found in the data by Monroy (2001).

Table 8. Stimuli for the sentence-reading task measuring controlled production of English /b d g/

	/b/				/d/				/g/		
	familiar		**novel**		**familiar**		**novel**		**familiar**		**novel**
1	about	**11**	*lobby*	**16**	today	**26**	*addict*	**31**	together	**41**	*sugar*
2	ability	**12**	*rabbit*	**17**	idea	**27**	*corridor*	**32**	ago	**42**	*tiger*
3	abuse	**13**	*global*	**18**	adore	**28**	*melody*	**33**	again	**43**	*legal*
4	*maybe*	**14**	*abbot*	**19**	*confident*	**29**	*modern*	**34**	*negative*	**44**	*yogurt*
5	*nobody*	**15**	above	**20**	*study*	**30**	adapt	**35**	*programme*	**45**	begin
6	bit			**21**	didn't			**36**	guide		
7	benefits			**22**	difficult			**37**	government		
8	BBC			**23**	day			**38**	good		
9	bottle			**24**	don't			**39**	guess		
10	be			**25**	dancer			**40**	going		

Note: Words featuring the sounds in unaccented position are represented in italics.

The second part of the production test was the timed picture-description task. Eight words were included in order to test the learners' spontaneous production of /z/ (five familiar, three novel), together with three distractors featuring /ʃ/. The three novel words were new in terms of spelling and position of the target sound within the word (see Table 9). However, the novel words included in this task were words that feature the sound /z/ under the spelling <z> in initial position. Since the sound is represented by a grapheme coinciding with the phonemic symbol, it was considered that generalising pronunciation gains to this somewhat 'prototypical' spelling should not be as hard as to others (e.g. *president*).

As regards the stimuli for English /b d g/, five tokens were selected for each sound (i.e. a total of 15 tokens). Given that this task was meant to elicit spontaneous productions of the target sounds, only words featuring the sound in intervocalic medial position were included. The sentence-reading task permitted the inclusion of words with the sound in initial position given that the researcher had already planned carrier sentences with the preceding words ending in a vowel. However, this time only medial positions were considered because learners' productions could not be controlled for. For example, even if a word like *bar* was preceded by determiners like *a* or *the*, if a learner mistakenly pronounced *an*, it would yield an unreliable measure of the /b/ in *bar* as a stop given that /b/ is also realised a stop after nasal sounds in Spanish. Tokens in this task also included familiar and novel words. Nevertheless, it was hypothesised that this should not affect the pronunciation of /b d g/ considerably, as once learners become aware of the rule, they should avoid spirantising these sounds in any position, regardless of their familiarity with the stimuli (see Table 9).

Table 9. Stimuli for the timed picture-description task for /z/ and /b d g/

	/z/ familiar		/b/ familiar		/d/ familiar		/g/ familiar		/ʃ/ distractors
1	music	**9**	about	**14**	idea	**19**	again	**24**	shower
2	example	**10**	maybe	**15**	study	**20**	negative	**25**	sheep
3	reason	**11**	nobody					**26**	fish
4	easy								
5	noise								
	novel		**novel**		**novel**		**novel**		
6	zoo	**12**	rabbit	**16**	addict	**21**	sugar		
7	zebra	**13**	global	**17**	melody	**22**	tiger		
8	zoom			**18**	modern	**23**	legal		

Finally, for the imitation task, five stimuli were selected featuring /z/ in initial (n=3), and medial (n=2) position. A further selection of 15 words was included in order to test English /b d g/ with the sounds in intervocalic position, five words per sound. As in the above tasks, three extra words with the voiceless post-alveolar fricative /ʃ/ were used as distractors (see Table 10).[87]

Table 10. Stimuli for the imitation task for /z/ and English /b d g/

	/z/		/b/		/d/		/g/		/ʃ/
1	zeal	**6**	abundance	**11**	comedy	**16**	rigour	**21**	sheep
2	zing	**7**	abbey	**12**	addict	**17**	forget	**22**	issue
3	z	**8**	ability	**13**	Adam	**18**	begin	**23**	fish
4	raisin	**9**	October	**14**	idiom	**19**	negotiate		
5	fuzzy	**10**	Iberia	**15**	adequate	**20**	figure		

g) Procedure for the evaluation of production stimuli

In order to analyse these data, the audios from the production pre- and post-tests were edited individually with Audacity (10 minutes approx. x 45 students[88] x pre and post-tests =

[87] Some of these words were also included in the perception tests. However, even though the identification task was conducted before, this was not considered to affect perception tests in any way, as learners never saw the words written in the identification task and they were presented amidst many other words and distractors.

[88] It must be noted that despite the fact that 47 participants took part in this study, only 45 are considered in the production test. As explained above, two of the post-test recordings had to be omitted from the analysis due to a very poor audio quality (participants 13 and 27), and therefore, the pre-test recordings were omitted too.

around 15.7 hours of audio). Each token featuring target items was saved as an individual file for its subsequent evaluation by the three expert judges described in section 3.3 above. The total number of items across tasks should yield a total of 11,070 tokens (Table 11).

Table 11. Number of tokens obtained per target item, task, in total, and for all participants (study 1)

Accent-imitation task				
/b d g/	/z/	Total	x2 (pre- and post-tests)	x 45 participants
15	-	15	30	1350
Sentence-reading task				
/b d g/	/z/	Total	x2 (pre- and post-tests)	x 45 participants
45	20	65	130	5850
Picture-description task				
/b d g/	/z/	Total	x2 (pre- and post-tests)	x 45 participants
15	8	23	46	2070
Imitation task				
/b d g/	/z/	Total	x2 (pre- and post-tests)	x 45 participants
15	5	20	40	1800
Total			**246**	**11070**
Total after missing audios				**9578**

Finally, as pointed out above, a total of 20 tokens were evaluated twice in order to calculate intra-rater reliability. Tokens included five tokens from each sound, with the same number of pre- and post-test productions, always by different speakers. For example, five words featuring /b/ in intervocalic position (*maybe, nobody, benefits, bottle*, and *be*), five random tokens by five random participants were selected for double evaluation, three from the pre-test and two from the post-test. This was a way of checking whether judges had assigned the same rating to those stimuli both times. Table 12 illustrates this arrangement visually.

Table 12. Sample of stimuli selection for intra-rater reliability for study 1

Target sound	Tokens	Participant	Pre- or post-test
/b/	maybe	12	pre
	nobody	18	pre
	benefits	19	post
	bottle	16	post
	be	11	post
/d/	today	5	pre
	adore	17	post
	study	15	post
	didn't	8	post
	dancer	7	pre
/g/	together	31	post
	ago	32	pre
	again	34	post
	negative	35	pre
	programme	36	post
/z/	music	40	pre
	reason	46	post
	example	42	pre
	residential	43	pre
	easy	44	pre
Summary	20 tokens	20 participants	10 pre/10 post

4.2.3 Results

As explained in section 3.1.2 above (see Figure 38 for a summary), the learners' perception of /s – z/ was measured with discrimination (task 1) and identification tasks (task 2). Perception of /b d g/ was measured by an accent mimicry task in which learners had to imitate English accented Spanish (task 3) and by an identification task that required learners to decide whether the degree of occlusion in Spanish and English voice stops was the same or not and (task 4). Finally, the learners' production of both target aspects was measured with three tasks, an imitation task (task 5), a sentence-reading task (task 6) and a timed picture-description task (task 7).

Given that both groups in this study acted as control and experimental, each group was expected to make improvements on different aspects. Group 1 (G1) received training on the English /s – z/ contrast and group 2 (G2) received instruction on English voiced stops. Thus, in perceptual tasks, it was expected that one improved their perception of the /s – z/ contrast (G1) and that the other improved their perception of occlusion in voiced stops /b d g/ (G2). Likewise, G1 was not expected to improve their perception of /b d g/ and G2 was not expected to improve their perception of the /s – z/ contrast. More specifically, G1 was expected to improve their scores in tasks 1 and 2, and G2 was expected to improve in tasks 3 and 4. With regard to production, the same tasks were used in order to assess both target aspects.[89]

Perception
Task 1. Discrimination task: /s – z/ contrast
Task 2. Identification task: /s – z/ contrast
Task 3. Accent mimicry task: occlusion of voiced stops /b d g/ in English accented Spanish
Task 4. Identification task: occlusion of voiced stops /b d g/
Production
Task 5. Imitation task
Task 6. Sentence-reading task
Task 7. Timed picture-description task

Figure 38. Tasks employed to measure the learners' perception and production of the target aspects in study 1

The results obtained by the two groups were analysed using two-way repeated measures ANOVAS, with group as between subjects factor and time (pre- and post-test) as within subjects factor. Given that ANOVA assumes data to be normally distributed, Shapiro-Wilk tests were conducted first in order to test the normality of data and Barlett and Levene tests were carried out in order to check whether the samples in these two groups have equal variances. The former were used when data were normally distributed, the latter when they were not.

Perception

The discrimination task (task 1) consisted of a total of 35 triads, including 10 triads aimed at familiarising the learners with the procedure and 5 distractors. Scores for pre- and post-tests were calculated without counting the learners' choices for distractors and training items. Thus, scores for the /s – z/ distinction were obtained by counting the number of correct responses out of 20 triads.

As explained in section 3.1.2, the discrimination tasks were also administered to three native speakers in order to test the validity of stimuli. If a triad was not discriminated correctly by native speakers, stimuli were analysed by the researcher and replaced with a

[89] Due to a problem with the computers at university, the pre-test data from four participants in the perception tests was lost and therefore neither pre- or post-tests are considered in the analysis of tests 1 and 2. They were participants 6 and 48 from G1 and participants 40 and 45 from G2.

more representative example of the sound (i.e. more prototypical). The results obtained by native speakers on the test were calculated by counting the number of correct responses, excluding distractors. The scores obtained by native speakers were 19 (95%), 17 (85%) and 20 (100%). Even though two of the native speakers did not obtain the maximum scores, their results were above 85%, suggesting that the test was a reliable measure of perceptual categories for the /s – z/ contrast. This is further supported by the scores obtained by the three expert judges employed for evaluation of stimuli. The non-native judges obtained similar scores than those by native speakers, sometimes even better. Their scores were 19 (95%), 19 and 18 (90%).

Regarding the participants' scores for this taks, the results for the discrimination task revealed that there was a significant effect of the time variable ($F(1,41) = 11.135$, $p = 0.002 < 0.05$), although no interaction was found between group and time ($F(1,41) = 0.27988$, $p = 0.59963 > 0.05$). In other words, there was a significant improvement from the pre-test to the post-test but there were not significant differences between groups. The mean scores for G1 were 9.6 (1.9) in the pre-test (48%) and 11.2 (2.4) in the post-test (56%), showing an improvement of 1.7 points (8.5%). The mean scores for G2 were 9.8 (3.7) in the pre-test (49%) and 11 (3.2) in the post-test (55%), with an improvement of 1.2 points (6%). Thus, both groups showed a similar improvement in their discrimination of the /s – z/ contrast (Figure 39). As Figure 39 shows, the mean scores for both groups were between nine and 12 points out of 20 at both testing times. That is, the mean scores by participants in both groups were always around the 50% performance.

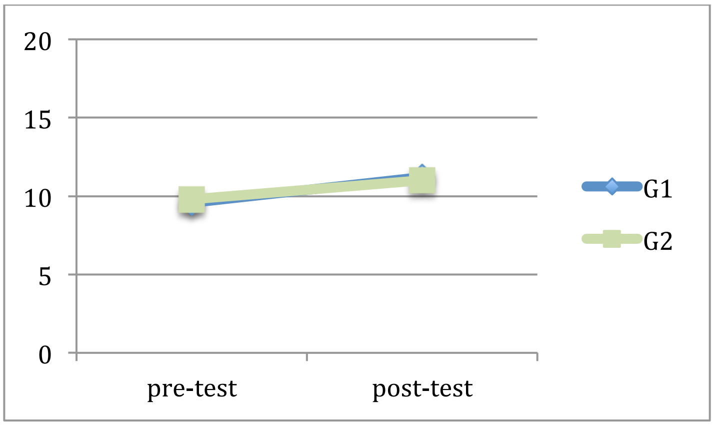


Figure 39. Mean scores for the /s – z/ contrast in the discrimination task at pre- and post-tests

Task 2 consisted of 40 stimuli aimed at measuring the learners' abilities to classify exemplars of /s/ and /z/ in an identification task. This task was also administered to the three non-native judges that evaluated stimuli from the production tasks. Their results were obtained by counting the number of correct responses in the 40 stimuli addressing the target contrast. The scores obtained by the three judges were 39 (97.5%), 37 (92.5%) and 38 (95%).

With regard to the participants' total scores in this task, the results show a significant effect of the time factor ($F(1,41) = 56.852$, $p = 2.878 \times 10^{-9} < 0.05$) and a significant interaction effect between time and group ($F(1,41) = 4.14$, $p = 0.048 < 0.05$). The mean scores in the pre-test were 25.5 (6.2) for G1 (63.8%) and 24.1 (7.2) for G2 (60.3%). The post-test scores were 31.5 (4.3) for G1 (78.8%) and 27.5 (6.6) for G2 (68.8%). The scores for the group receiving instruction (G1) increased by six points (15%) from pre- to post-test, while the ones for the control group (G2) increased by 3.4 points (8.5%). That is, there were significant improvements from pre- to post-test, with G1 showing a significantly higher improvement than the one by G2 (Figure 40).

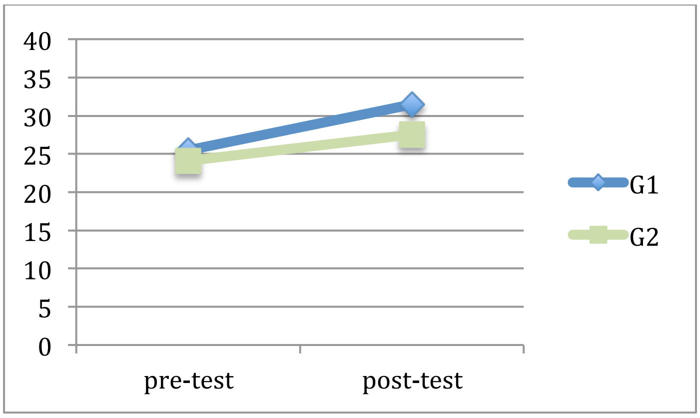


Figure 40. Mean scores for the /s – z/ contrast in the identification task at pre- and post-tests

The fact that learners in G1 improved their mean scores for the total number of items (including all instantiations of /s/ and /z/) indicates that their ability to identify instantiations of these sounds increased. However, given the 'I don't know' option in the identification test, it was important to explore whether there were differences in the identification scores for both sounds. It might be the case that learners were more capable of identifying one sound than the other. As a case in point, learners might be able to recognise examples of /s/ as /s/ clearly, given the presence of a similar phoneme in their L1, but hesitate when deciding whether a stimulus was an example or /z/ or not. Thus, in order to investigate possible differences between sounds, two ANOVAs were performed dealing with the scores obtained in items featuring /s/ and items featuring /z/ separately.

The mean scores obtained by both groups show that G1 outperformed G2 for both sounds (Table 13). However, despite tha fact that the results for /s/ revealed a significant effect of the time variable ($F(1,41) = 13.168$, $p = 0.001 < 0.05$), no interaction was found between groups ($F(1,41) = 3.024$, $p = 0.0895 > 0.05$). That is, there was a significant improvement from pre- to post-test in the learners' perception of /s/, but the difference between groups was not significant. The results for /z/ yielded a very similar result, with a significant effect of time ($F(1,41) = 35.264$, $p = 5.295 \times 10^{-7} < 0.05$) but no interaction between groups ($F(1,41) = 1.117$, $p = 0.296 > 0.05$). The biggest improvements occurred in the identification of /z/, with an increase of 17.8 per cent in the scores by G1 and 12.3 per cent in the ones by G2. Also, it is worth noting that while both groups' scores were around 50 per cent of the total score for /z/ in the pre-test, they were near or over 70 per cent for /s/.

Table 13. Mean scores (with standard deviations in parentheses) and improvement (imp) made in the identification task considering the scores for /s/ and /z/ separately

	/s/					
	pre-test	**%**	**pos-test**	**%**	**imp**	**%**
G1	14.4 (2.5)	72.2	16.6 (2.3)	82.8	2.0	9.8
G2	13.7 (3.4)	68.3	14.4 (3.6)	72	0.7	3.4
	/z/					
	pre-test	**%**	**pos-test**	**%**	**imp**	**%**
G1	11.0 (4.9)	55.2	14.9 (4.0)	74.6	3.6	17.8
G2	10.5 (4.4)	52.2	13.2 (3.8)	65.8	2.5	12.3

Additionally, as explained in section 4.2.2f, testing stimuli were divided into familiar and novel words in order to test possible generalisation gains. In order to reflect robust category formation learners should be able to identify exemplars of these sounds even if they occur in words that have not appeared in training. As explained above, there were 20 familiar and 20 novel words in order to test the learners' identification of /s/ and /z/ (10 for each sound). Table 14 shows the mean scores for familiar and novel stimuli obtained by both groups at pre- and post-tests including the improvement made between testing times.

Table 14. Mean scores and improvement (imp) in identification of /s/ and /z/ for familiar and novel items in task 2

	familiar						**novel**					
	pre	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
G1	12.3 (3.3)	61.3	15.8 (2.2)	79.1	3.6	17.8	13.2 (3.3)	66.1	15.7 (2.5)	78.3	2.4	12.2
G2	11.4 (3.8)	57	14.0 (3.4)	69.8	2.6	12.8	12.7 (3.9)	63.5	13.6 (3.7)	68	0.9	4.5

The ANOVA comparing both groups' scores for novel stimuli reveal a significant effect of the time variable ($F(1,41) = 15.2$, $p = 0.000355 < 0.05$) but no significant interaction effects between time and group ($F(1,41) = 3.02435$, $p = 0.08953 > 0.05$). That is, both groups showed a similar improvement from pre- to post-test, although the average improvement made by G1 was one point higher than the one by G2 in familiar items and 1.5 points higher in novel items.

Focusing on the scores for /s/, a significant interaction effect between time and group was found for novel stimuli ($F(1,41) = 6.66$, $p = 0.0136 < 0.05$). The mean scores for G1 in novel words increased from 6.9 in the pre-test to 8.3 in the post-test (an increase of 1.4 points), whereas the control group's performance was the same in the pre- and post-test, with 6.8 points at both testing times (Table 15).

As regards improvements in the perception of /z/, the ANOVA revealed a significant effect of the time variable ($F(1,41) = 9.91$, $p = 0.00306 < 0.05$), but no interaction between time and group ($F(1,41) = 0.02$, $p = 0.869 > 0.05$). That is, even though the improvements made by G1 were higher than those by G2, these did not reach significance. It is worth noting that both groups' scores for /z/ were higher with novel than with familiar stimuli. The scores for /z/ with familiar items in the pre-test were lower than those for /s/. However, the scores for both sounds in novel stimuli were similar.

Table 15. Mean scores and degree of improvement (imp) in familiar and novel words for /s/ and /z/ in the identification task

	/s/											
	Familiar						**Novel**					
	pre	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
G1	7.6 (1.4)	75.7	8.3 (1.4)	82.6	0.7	7	6.9 (1.6)	68.7	8.3 (1.3)	83	1.4	14
G2	6.9 (1.7)	68.5	7.6 (1.8)	76	0.8	8	6.8 (2)	68	6.8 (2.2)	68	0	-
	/z/											
	Familiar						**Novel**					
	pre	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
G1	4.7 (2.8)	47	7.6 (2.1)	75.7	2.9	29	6.3 (2.3)	63.5	7.3 (2.1)	73.5	1	10
G2	4.6 (2.5)	45.5	6.4 (2.2)	63.5	1.8	18	5.9 (2.1)	59	6.8 (1.9)	68	0.9	9

As explained above, the learners' perception of /b d g/ as stops in intervocalic position was assessed by means of two tasks (see section 3.1.2). The first was a delayed accent-mimicry task (task 3). The second was an identification task in which learners compared examples of /b d g/ in English and Spanish words and decided whether their degree of occlusion was the same or different (task 4).

As mentioned in section 3.2, the delayed accent-mimicry task (task 3) was conducted as part of the production test and was evaluated by the same three non-native judges. The judges' ratings were dichotomous (i.e. 1 for an acceptable occlusion in the pronunciation of stops or 0 for non-acceptable occlusion). Once judges had completed their evaluations, the three sets of ratings were compared and judges met again in order to disambiguate disagreements. In those cases in which judges could not reach an agreement, spectrographic analyses of learners' productions of /b d g/ were performed. The final ratings for this test were saved on a separate Excel spreadsheet. However, the original scores given by the three judges were saved in order to check inter-rater reliability. The results of the Fleiss' Kappa test reveal an interrater reliability of 0.941, which can be interpreted as 'almost perfect agreement' (range 0.81-1.00).

As explained in section 4.2.2f, the delayed accent-mimicry task included 15 words featuring /b d g/ (5 tokens for each sound) together with 15 distractors featuring /p t k/ (5 tokens each). Table 16 shows the mean scores for G1 and G2 in the pre- and post-tests, including the degree of improvement. The table includes the scores obtained for /b d g/ as a whole (maximum score = 15 points) as well as the data for /b/, /d/ and /g/ separately (maximum score = 5). G1 decreased their average performance for /b d g/ by 0.4 points from pre- to post-test, indicating that no gains were obtained for the group that did not receive instruction. On the contrary, learners in G2 improved their average performance for /b d g/ by 0.8 points, starting training with 8.1 out of 15 and obtaining 9 points after receiving training. Nonetheless, as revealed by the ANOVAs performed for these variables, this improvement did not reach significance.

Table 16. Mean scores and SD at pre- and post-tests and degree of improvement (imp) in task 3

	/b/			**/d/**			**/g/**			**/bdg/**		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	2.8 (1.6)	2.7 (1.7)	-0.1	3.9 (1.3)	3.6 (1.6)	-0.3	3.2 (1.6)	3.2 (1.5)	0	9.9 (4)	9.5 (4.1)	-0.4
%	56.7	54.2	-2.5	77.5	72.5	-5	62.2	63.3	0	66.1	63.3	-2.7
G2	2.1 (1.7)	2.8 (1.9)	0.6	3.4 (1.6)	3.5 (1.6)	0.1	2.6 (1.6)	2.7 (1.6)	0.1	8.1 (4.2)	9.0 (4)	0.8
%	42.9	55.2	12.4	67.6	70.5	2.9	52.4	53.3	1.0	54.3	59.7	5.4

Table 16 shows that participants in the control group (G1) made no improvements from pre- to post-test, even showing decreases in their post-test scores. Participants in G2 made the biggest improvement in /b/ (12.4% of improvement), while increases in the mean scores for /d/ and /g/ were very moderate (2.9%).

The last perception task was aimed at testing the learners' ability to perceive differences in occlusion between English and Spanish /b d g/ (task 4). It consisted of 48 items with nine distractors featuring /p t k/. However, since distractors are not considered in the analysis, the total score in this test was 39 points, with 13 items featuring each of the three target sounds. This task was also administered to two of the non-native judges that evaluated stimuli from the production tasks. The researcher did not perform this task as it was him who had designed the test and it would not provide a reliable measure of the test's effectiveness. Judges were expert phoneticians and might be aware of the fact that /b d g/ were spirantised in Spanish when flanked by vowels but not after nasal sounds. However, they were told that stimuli were a random selection of acoustically altered tokens and that they sould rely solely on their perception of the sound as sharing the occlusion in both languages or not. The two judges' scores were 30 out of 30 (i.e. 100%).

The results obtained from the ANOVA comparing learners' mean scores at pre- and post-tests for the three sounds (i.e. out of 39 points) revealed that despite subtle differences between groups, these did not reach significance ($F(1,41) = 0.770$, $p = 0.385 > 0.05$). G1 had a mean score of 24.7 (4.5) at the beginning of the study (63.3%) and this score decreased to 24.1 (3.8) in the post-test (61.8%). G2, on the other hand, started with a mean score of 24.9 (3.6), which represented 63.8 per cent of the total score, and finished the study with a mean score of 25.5 (5.6), 65.4 per cent of the total for this task. The mean scores for G1 decreased by 0.6 points (1.5%) whereas those for G2 increased in the same 0.6 points. These data reveal that both groups started with very similar identification scores at the beginning of the study, but G2 was the only group that made improvements from pre- to post-test – albeit very moderate (see Figure 41). However, it should be noted that the learners' scores for both groups were already high at the beginning of the study, as they started with a mean score of 24.8 (4.07) points out of 39, which stands for 63.5% of the total score they could get.

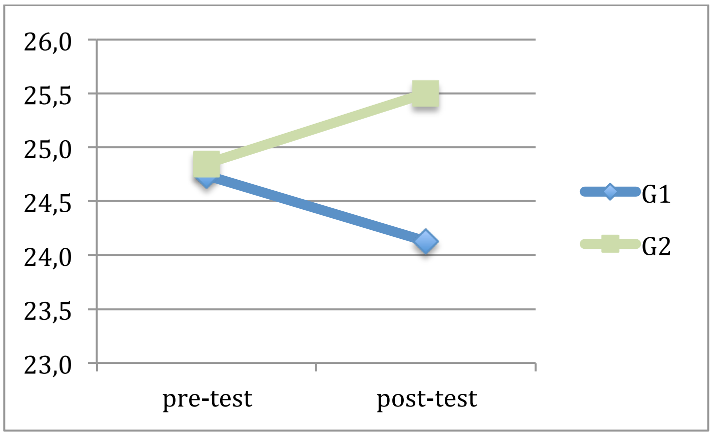


Figure 41. Mean scores for /b d g/ in task 4 in the pre- and post-tests

In order to investigate possible differences among the three target sounds, the mean scores for both groups were calculated by separating the scores obtained in items featuring each target sound. The results from the ANOVAs revealed that there were no significant effects of the time variable and no interaction effects of the group variable for any of the sounds.

Table 17 shows the mean scores obtained by each group for each of the three sounds in the pre- and post-tests, including the degree of improvement. The maximum score for each sound was 13 points, as there were 13 items featuring each sound. The average scores obtained for both groups for each sound reveal that the learners' performance was very similar across sounds. The mean scores for G1 were distributed between 7.7 and 8.7 points (i.e. within a one point range) and data for G2 were distributed between 8.1 and 8.8 points (i.e. within a 0.7 range). The scores for both groups were all within the 7.7 and the 8.8 range, suggesting that the participants' ability to perceive differences in occlusion across these three sounds was very similar. Even though G2 outperformed G1, the differences were minimal. Furthermore, it should be noted that both groups' performance in this task was already adequate in the pre-test, as both group's mean scores were already above 60 per cent of the maximum score they could get in this task.

Table 17. Average scores for /b/, /d/ and /g/ in task 4 at pre- and post-tests including the degree of improvement (imp)

	/b/			/d/			/g/		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	8.4 (1.7)	7.7 (2)	-0.7	7.8 (1.9)	7.7 (1.9)	-0.1	8.5 (2.3)	8.7 (1.9)	0.2
%	64.6	59.2	-5.4	60	59.2	-7.7	65.4	66.9	1.5
G2	8.5 (1.5)	8.8 (2.4)	0.3	8.1 (1.8)	8.3 (1.9)	0.3	8.3 (2.6)	8.5 (2.5)	0.1
%	65.4	67.7	2.3	62.3	63.8	2.3	63.8	65.4	0.8

Production

With regard to production, there were three different tasks aimed at measuring the learners' performance on the target aspects, namely an imitation task (task 5), a sentence-reading task (task 6), and a timed picture-description task (task 7) – see section 3.1.5. As with the perception data, the results are presented by considering the total scores for each target aspect first, and then focusing on the specific results obtained in each of the three tasks. Table 18 shows the maximum score for each sound in these tasks, including familiar and novel stimuli.[90]

Table 18. Maximum score for each sound in tasks 5, 6 and 7

Stimuli	Task 5	Task 6	Task 7	Total
/z/	5	20 (10 familiar/10 novel)	8	33
/b/	5	15 (10 familiar/5 novel)	5	25
/d/	5	15 (10 familiar/5 novel)	5	25
/g/	5	15 (10 familiar/5 novel)	5	25
/b d g/	15	45	15	75

It is important to point out that, even though the total number of tokens should be 11,070 (see section 4.2.2g above), some learners occasionally skipped words in one of the tests (pre or post). Consequently, both pre- and post-test recordings for that token were omitted from the analysis and are not taken into consideration for the results. This was sometimes due to the fact that learners pressed the key too fast when handling the PowerPoint and skipped one or several slides, or simply because they chose not to pronounce the word. The final number of tokens for evaluation was 9,578.

As in the delayed accent-mimicry task, interrater reliability was measured using Fleiss' Kappa test. The interrater reliability was found to be 0.94, which can be interpreted as 'almost perfect agreement' (0.81-1.00 range). Furthermore, as explained in section 3.3,

[90] The post-test recordings by two participants (13 and 27) could not be evaluated due to a very poor audio quality. Thus, data from these participants in the production tasks are not taken into account in the analysis.

intra-rater reliability, was measured by playing a selection of 20 extra items that had already been assessed. In this study, there was only one item that did not receive the same score by one of the judges. Given this, no extra tests were conducted, as intra-rater reliability was considered to show an almost perfect agreement too.

Given that the judges' decisions when evaluating learners' productions were dichotomous (i.e. 1 or 0), the total scores obtained for each sound were calculated by counting the number of correct productions for each target sound across tasks at both testing times and then calculating the difference between pre-tests and post-tests (i.e. the degree of improvement). Table 19 shows the mean scores for each of the target sounds in pre- and post-tests for G1 and G2, including the degree of improvement (imp) and the maximum score for each sound (Max). The results show that despite moderate improvements in the pronunciation of /b d g/ as stops by G1 (the group receiving instruction on the /s – z/ contrast), it was G2 that obtained the largest improvement.

Table 19. Mean scores, standard deviation and improvement in the participants' production of each target sound across tasks

	/b/			/d/			/g/			/z/		
Max	25			25			25			33		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	13.4 (5.8)	15.0 (5.6)	1.6	18.3 (4.8)	19.0 (5.3)	0.6	14.4 (4.9)	14.8 (5.2)	0.4	8.5 (8)	12.5 (4.7)	4.0
%	53.5	60	6.5	73.3	75.8	2.5	57.7	59.2	1.5	25.8	37.8	12
G2	11.8 (5.4)	15.1 (4.8)	3.3	17.1 (4.2)	18.7 (3.1)	1.5	13.0 (5.7)	15.1 (4.4)	2.2	3.4 (5)	4.7 (5.3)	1.3
%	47.2	60.6	13.3	68.6	74.7	6.1	51.8	60.6	8.8	10.4	14.3	3.9

The production data were also analysed using mixed ANOVAs, with group as between subjects factor and time as within subjects factor. The ANOVA conducted for /z/ revealed a significant effect of group ($F(1, 43) = 9.71$, $p = 0.00326 < 0.05$) and time ($F(1, 43) = 28.6$, $p = 3.21\text{e-}06 < 0.05$), and an interaction effect between group and time ($F(1, 43) = 7.43$, $p = 0.00926 < 0.05$). This indicates that there were significantly different scores between groups at all times, but that regardless of these differences, one of the groups made significantly higher improvements than the other group (Figure 42). Participants in G1 started with a mean score of 8.5 points (25.8%) and finished with a mean score of 12.5 (37.8%). That is, they improved by 4 points (12%) from pre- to post-test. The scores by participants in G2 were comparatively lower, with a mean score of 3.4 points (10.4%) in the pre-test and 4.7 (14.3%), improving by 1.3 points (3.9%) from pre- to post-test.

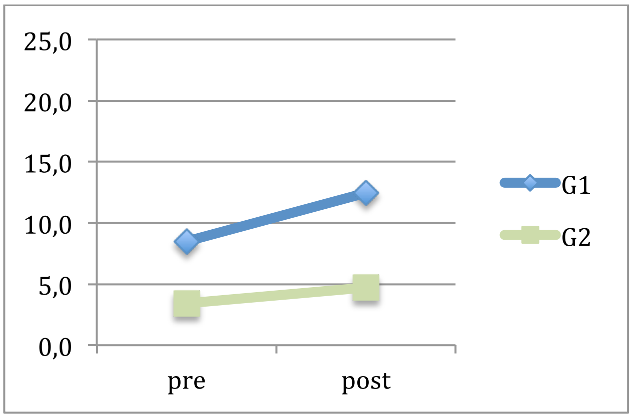


Figure 42. Pre- and post-test scores for G1 and G2 on the production of /z/

Focusing on the results obtained for /b d g/ as a whole (i.e. out of a maximum of 75 points), the mean scores for G2 improved by 7 points from pre- to post-test (9.4%), whereas the scores for G1 improved by 2.6 points (3.5%). The results reveal a significant effect of time and an interaction between time and group ($F(1, 43) = 5.33$, $p = 0.0259 < 0.05$). That is, even though both groups' scores for /b d g/ increased from pre- to post-test, there was a significant difference between the performance of G2 and that of G1 (Figure 43).

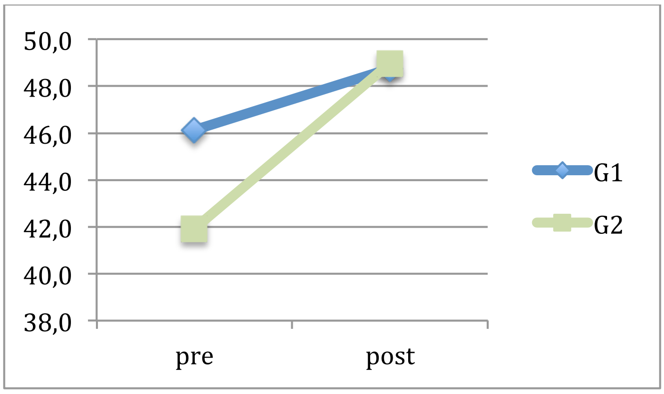


Figure 43. Pre- and post-test scores on the production of /b d g/ across tasks

Furthermore, a closer inspection of the data obtained for /b d g/ for G2 reveals that the sound which improved most was /b/, followed by /g/, and finally /d/ (Figure 44). However, while the overall improvement made by G2 for /b d g/ was considerably higher than that by G1, and despite the fact that G2 outperformed G1 in their scores for every sound, inter-group comparisons for individual sounds revealed that the only sound for which differences between groups reached significance was /g/ ($F(1, 43) = 5.53$, $p = 0.0233 < 0.05$).

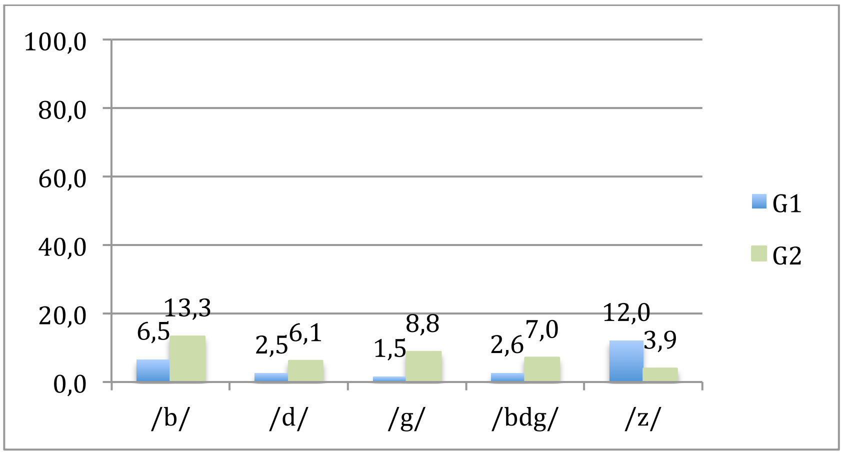


Figure 44. Percentage of improvement for each sound across production tasks

Focusing on the imitation task, the data reveal that the improvements made by both groups were very moderate (see Table 20). A significant difference was found for /z/ in the time variable ($F(1,43) = 11.6$, $p = 0.00143 < 0.05$) and in the group variable ($F(1,43) = 11.6$, $p = 0.00143 < 0.05$), but no interaction was found between time and group ($F(1,43) = 0.718$, $p = 0.401 > 0.05$). This implies that the improvement from pre- to post-test was considered significant for both groups and that there was a significant difference between these two groups' scores at both testing times. However, the differences in the improvement between groups were not significant. G2's scores for /z/ were considerably lower than those by G1 (Figure 45). The mean scores for G1 show an improvement of 0.5 points (9.2%) in their imitation of /z/. However, the average scores for G2 for the pronunciation of /z/ increased by 0.8 points (15.2%), showing a greater improvement than the one obtained by the group receiving instruction.

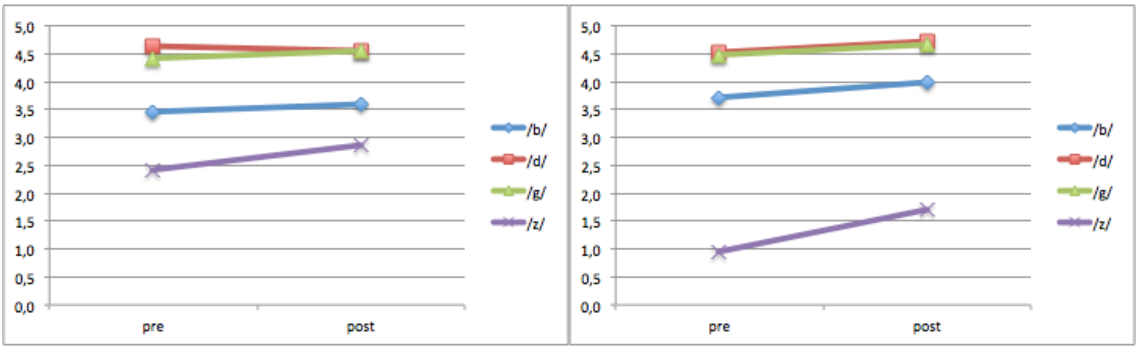


Figure 45. Mean scores for G1 (left) and G2 (right) in the imitation task at pre- and post-tests

With regard to /b d g/, G2 shows only slight improvements in their ability to imitate the pronunciation of /b d g/ as stops in intervocalic position. Taken as a whole, the learners' pronunciation of /b d g/ improved by 0.7 points (4.4%) from pre- to post-test, which is 0.5 points higher than the improvement obtained by G1, the group acting as control. Nonetheless, the improvement made for individual sounds was very moderate, with 0.3 points for /b/ (5.7%) and 0.2 points for /d/ and /g/ (3.8%), although it should be noted that both groups' mean scores showed ceiling effects for /b d g/ in the pre-test (Figure 45).

Table 20. Mean scores at pre- and post-test for the three production tasks including the degree of improvement made and the maximum (Max) score for each sound on each task

		G1						G2					
		Imitation task											
	Max	**pre**	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
/b/	5	3.5	69.2	3.6	71.7	0.1	2.5	3.7	74.3	4.0	80.0	0.3	5.7
/d/	5	4.6	92.5	4.5	90.8	-0.1	-1.7	4.5	90.5	4.7	94.3	0.2	3.8
/g/	5	4.4	88.3	4.5	90.8	0.1	2.5	4.5	89.5	4.7	93.3	0.2	3.8
/z/	5	2.4	48.3	2.9	57.5	0.5	9.2	1.0	19.0	1.7	34.3	0.8	15.2
		Sentence-reading task											
	Max	**pre**	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
/b/	15	8.2	54.7	9.3	62.2	1.1	7.5	7.2	47.9	9.6	64.1	2.4	16.2
/d/	15	10.8	71.7	11.2	74.4	0.4	2.8	9.7	64.8	10.8	71.7	1.0	7.0
/g/	15	8.4	56.1	8.5	56.9	0.1	0.8	7.2	48.3	9.1	60.6	1.9	12.4
/z/	20	4.7	23.3	7.6	38.1	3.0	14.8	2.1	10.7	2.4	11.9	0.2	1.2
		Timed picture-description task											
	Max	**pre**	**%**	**post**	**%**	**imp**	**%**	**pre**	**%**	**post**	**%**	**imp**	**%**
/b/	5	1.7	34.2	2.1	41.7	0.4	7.5	0.9	18.1	1.5	30.5	0.6	12.4
/d/	5	3.0	59.2	3.3	65.0	0.3	5.8	2.9	58.1	3.2	63.8	0.3	5.7
/g/	5	1.6	31.7	1.7	34.2	0.1	2.5	1.2	24.8	1.4	27.6	0.1	2.9
/z/	8	1.4	17.7	2.0	24.5	0.5	6.8	0.3	4.2	0.6	7.7	0.3	3.6

As for the sentence-reading task, the maximum scores for /b d g/ were 15 points for each sound (10 words featuring the sounds in familiar words and 5 in novel ones) and for /z/ 20 points (10 familiar words and 10 novel ones).

An ANOVA comparing both groups' mean scores for /z/ (including familiar and novel words) revealed significant effects of group ($F(1, 43) = 9.57$, $p = 0.00347 < 0.05$) and time ($F(1, 43) = 23.1$, $p = 0.0000189 < 0.05$) and an interaction effect between these two factors ($F(1, 43) = 16.7$, $p = 0.000184 < 0.05$). That is, there were significant differences in the mean scores of both groups at both testing times, but also in the degree of improvement made by each group. G1 showed a considerably higher improvement in /z/ than that by G2, which remained stable (Figure 46). G1 improved their pronunciation of /z/ in 2.96 points (14.8%) as opposed to G2 (acting as control), whose scores remained stable with a slight improvement of 0.2 points (1.2%).

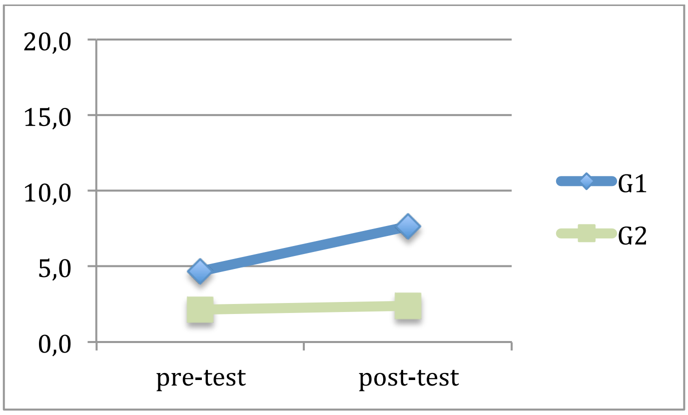


Figure 46. Mean scores for the production of /z/ in the sentence-reading task

As for /b d g/, the results for the three sounds taken together reveal a significant interaction effect between time and group ($F(1, 43) = 5.16$, $p = 0.0282 < 0.05$). G1 made an improvement of 1.7 points (3.7%) and G2 improved 5.3 points (11.9%). This indicates a significant effect of treatment for the group receiving instruction on /b d g/. However, when the data for /b d g/ were analysed individually (Figure 47), even though G2 made greater improvements for /b/ and /d/ than G1, /g/ was the only sound for which differences reached significance G1 ($F(1, 43) = 7.8$, $p = 0.00776 < 0.05$). Moreover, given that the maximum scores in this task were 15 points for /b d g/ and 20 for /z/, learners were still far from the maximum rating in this task.

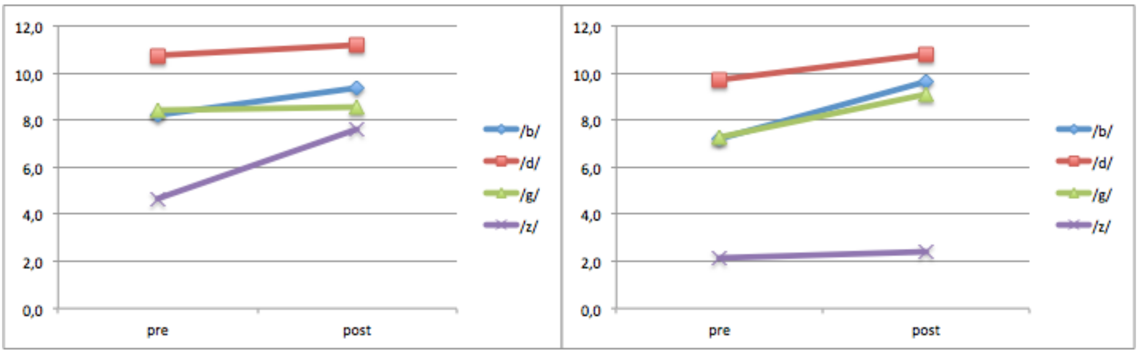


Figure 47. Mean scores for G1 (left) and G2 (right) on /b d g/ and /z/ in the sentence-reading task

In addition, as explained above, stimuli in this task were divided into familiar and novel words in order to test possible generalisation gains. For /b d g/, there were 10 familiar and 5 novel words for each sound, whereas /z/ was featured in 20 items (10 familiar and 10 novel).

An analysis of the learners' performance in familiar contexts revealed significant interaction effects of the group and time variables for the improvements made in /b/ ($F(1,43) = 4.33$, $p = 0.0435 < 0.05$), /g/ ($F(1,43) = 11.3$, $p = 0.00162 < 0.05$), and /z/

($F(1,43) = 10.5$, $p = 0.00228 < 0.05$). The improvement made by G1 on the pronunciation of /z/ was significantly higher than the one made by participants in G2. On the contrary, the improvements by participants in G2 on /b/ and /g/ were significantly higher than the ones made by participants in G1.

Focusing on the results for novel stimuli, the results for /z/ reveal a significant effect of group ($F(1, 43) = 9.57$, $p = 0.00347$) and time ($F(1, 43) = 23.1$, $p = 0.0000189 < 0.05$) and an interaction effect of time and group ($F(1, 43) = 16.7$, $p = 0.000184 < 0.05$). This indicates that there were significant differences between both groups' scores at both points in time. G2's scores for novel stimuli featuring /z/ were considerably lower than those by G1. In addition, while G2 scores remained stable from pre- to post-test, G1's scores improved significantly (Figure 47).

As for the improvements made in the pronunciation of /b d g/, a significant effect of the time variable was found when comparing both groups' performances on novel stimuli for /b d g/ across sounds ($F(1,45) = 6.05$, $p = 0.0178 < 0.05$). However, there were no significant interaction effects between group and time ($F(1,45) = 0.914$, $p = 0.344 > 0.05$). That is, both groups improved similarly on the overall scores for /b d g/. Furthermore, an analysis of the scores obtained in novel stimuli for individual sounds revealed that there were no significant differences between groups in the improvement of any of the three voiced stops. The increases in G2's scores are always larger than those of G1 except for /b/ in novel words, where participants in the control group outperformed those in the experimental group, albeit minimally. For all the other sounds, the improvement achieved by participants in G2 was greater than the one by participants in G1. With the exception of /b/, the improvement in novel words by participants in G1 for /d/ and /g/ was very moderate or non-existent (Figure 48).

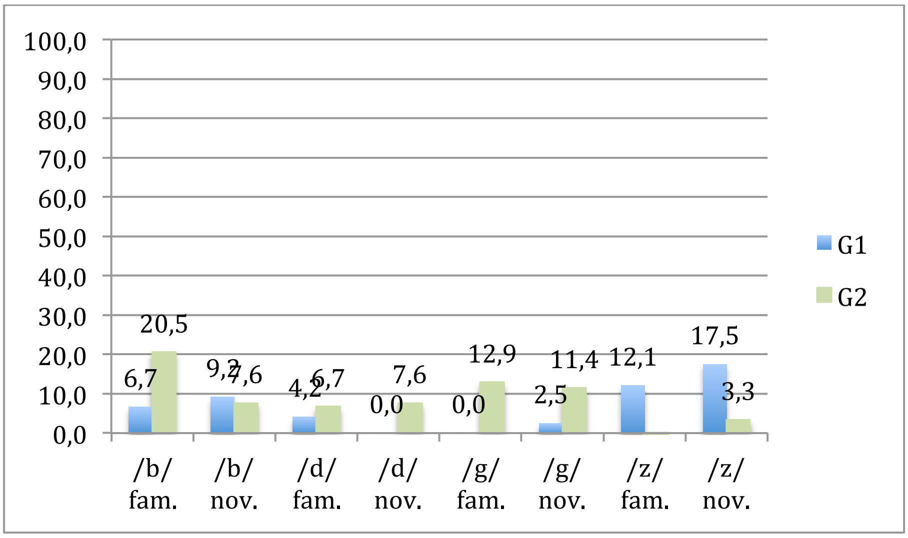


Figure 48. Percentage of improvement made by participants in G1 and G2 in familiar (fam.) and novel (nov.) words in the sentence-reading task

The last production task was the timed picture-description task (task 7), which measures the learners' ability to produce the target sounds in spontaneous production. As explained in section 4.2.2f, /b d g/ were featured in five words in this task (hence their maximum score was five), whereas /z/ was featured in eight (i.e. eight was the maximum score). What participants had to learn about /b d g/ was considered to be easily transferrable to new contexts, as these sounds are always represented by the same graphemes in spellings, furthermore coinciding with their phonemic symbols. Hence, stimuli featured the learners' ability to articulate this sound spontaneously both in familiar and novel words. As for /z/, three novel words were also included in order to test the participants' production of this sound in novel contexts represented by <z>, a very common spelling for this sound – also coinciding with its phonemic symbol.

Table 20 shows the average scores obtained in this task without considering whether words were familiar or novel. The results show that the improvement obtained by both groups from pre- to post-test was rather moderate in spontaneous production. Moreover, both groups showed similar improvements in both training aspects. No significant interactions were found between time and group for any of the sounds. G1 improved almost the same as G2 in their production of /b d g/ with the only exception of /b/, in which G2 showed a slightly higher improvement. Likewise, the scores for /z/ were only slightly lower for G2 than for G1 (0.2 points), the group receiving instruction.

Figure 49 illustrates the degree of improvement made for each target sound treating familiar and novel words separately. However, these data are only included here anecdotally, as the range of improvement obtained was extremely limited (between 0 and

0.5 points) and there were very few words to establish any generalisable comparisons. The data show that differences in the degree of improvement obtained by G1 and G2 were generally higher for words that appeared in training. As a case in point, the improvement obtained by participants in G2 for /b/ and /d/ was higher than that for participants in G1 (the control group), although the difference was rather subtle in the case of /d/ – similar to the results obtained in familiar items for /z/ by G1 and G2. However, in the case of novel words, differences between groups were extremely subtle for /b d g/, with both groups obtaining almost the same results – although for /z/, there is a clear difference in the improvement obtained by G1 (the one receiving training on /z/) and the results obtained by G2.

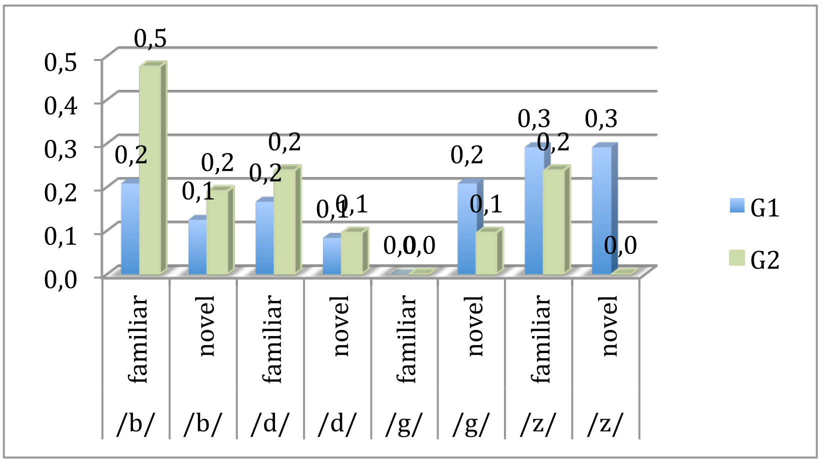


Figure 49. Improvement made in novel and familiar words in the timed picture-description task

Questionnaires

The above paragraphs have presented the empirical data obtained from the perception and production tests. The remainder of the section will address the quantitative and qualitative data obtained from the questionnaires administered before and after the pre- and post-tests respectively (Appendices 11 and 12).

In order to obtain a reliable measure of their perceptions towards the podcast-based instruction, participants were not required to fill in their participant number in the post-test questionnaires. This limits the reliability of pre-/post-test comparisons, since the learners' responses from the pre-test are linked to their participant number but their post-test responses are not. Furthermore, the number of participants who completed both interviews is not the same, as there were five less participants in the post-test. Therefore, despite possible changes from pre- and post-test scores in some items, these are only included anecdotally, as no comparisons can be made with a different number of respondents at each testing time.

Forty-two participants completed the pre-test questionnaires (24 from G1 and 18 from G2) and 37 completed the post-test questionnaires (20 from G1 and 17 from G2). Since learners in both groups received the same type of instruction, responses from participants in both groups will be combined. A number of 5-point Likert-type items were included in both questionnaires in order to investigate possible changes as a result of training (Table 21).

The podcast-based instruction was expected to help learners feel more comfortable when speaking in public and recording their voice. The participants' mean responses to items 1 and 2 in Table 21 show a slight decrease from pre- to post-test, albeit very moderate. However, the mean scores for items 1 and 2 before the pre-test suggest that participants did not feel particularly shy about speaking in public or recording their voice at the beginning of the study.

Given that the learners' attention was explicitly directed to form in activities 1 and 2 in Edmodo, and since they were required to evaluate their peers' pronunciation, it was expected that instruction helped them pay attention to the way people speak and develop self-monitoring skills to analyse and repair their own mistakes. Nonetheless, their responses to items 3 to 6 in Table 21 show little or no improvements from pre- to post-test.

Table 21. Mean scores in the Likert-type questions from pre- and post-tests questionnaires (study 1)

	Item	Pre (n = 42)	Post (n = 37)
1.	I feel shy about speaking English in public	3.1 (1.2)	3 (1.4)
2.	I feel shy about recording my voice if other people are going to hear the recording	2.8 (1.2)	2.4 (1.3)
3.	Do you pay attention to the way people pronounce English and compare your pronunciation with theirs?	4.2 (0.7)	4 (0.7)
4.	Do you realise when you make pronunciation mistakes?	3.8 (0.5)	3.8 (0.6)
5.	When you make pronunciation mistakes, do you realise what the problem is?	3.5 (0.7)	3.8 (0.6)
6.	When you make a pronunciation mistake, are you able to correct the mistake yourself?	3.4 (0.6)	3.5 (0.6)

Post-test questionnaires also included a number of items using Likert-type scales in order to obtain information regarding the learners' opinions about the approach adopted and their use of podcasts during those weeks (Table 22). The learners' responses suggest that they considered the length of the study to be adequate in order to improve their pronunciation of the target aspects (item 7). The mean scores for item 11 indicate that they do not think less time would have been as effective in order to make improvements, and their responses to item 12 suggest that some of them would have appreciated some more practice. Despite the different tasks they had to complete weekly, overall, they did not consider the study excessively demanding (item 18).

Items addressing the learners' perceived usefulness of the approach show the highest mean scores. This indicates that participants considered the method had helped them improve their perception and production of the target sounds (items 8 and 9) and that they found it potentially advantageous in order to tackle problematic pronunciation aspects (item 10). Furthermore, their responses suggest that the aspects covered were perceived as being close to their needs and applicable to their daily use of English (item 17).

The online perceptual activities carried out in Edmodo were considered to be beneficial in order to help learners perceive the target sounds better (item 13). In fact, when asked to choose the activity they found most useful, 10 students chose activity 1 (looking for instantiations of the target sounds in the podcasts), 14 chose activity 2 (the multiple-choice sound-identification activity) and 13 chose activity 3 (recording their own podcasts). That is, 24 students chose the perceptual activities over recording their own podcasts. This does not mean that they did not perceive recordings to be useful, but perhaps they considered some kind of prior instruction necessary in order to make the most of the recordings. The participants' responses show that they value the feedback received by their peers (items 14 and 15) and that they felt capable of offering feedback themselves (item 16), although activities 1 and 2 were also valued higher than this, as indicated in the mean response to item 13.

Table 22. Items from post-test questionnaires with mean responses and standard deviation in brackets

Item	Mean (n = 37)
7. I consider the length of the study adequate so as to improve my pronunciation of the sounds covered	3.9 (1)
8. The method employed has helped me perceive the target sounds better	4.5 (0.6)
9. The method employed has helped me improve my pronunciation of the target sounds	4.5 (0.6)
10. I find the method to be helpful in order to work on problematic pronunciation aspects	4.4 (0.7)
11. Less time would have been equally effective in order to help us improve the sounds addressed	2 (1.1)
12. I still need more practice in order to master those sounds. We have not had enough time.	2.8 (1.1)
13. The weekly online activities in Edmodo (activities 1 and 2) have helped me improve my perception of the target sounds	4.4 (0.6)
14. Feedback from my peers has been adequate. Their assessments were correct.	3.8 (0.8)
15. The peer feedback received has helped me improve.	3.7 (1)
16. I have felt capable of offering feedback to my peers	4 (0.8)
17. I find the aspects covered helpful for my daily use of English	4.4 (0.7)
18. I have found the study too demanding	1.8 (0.6)

Additionally, they were asked about the degree of involvement they perceived in their peers' recordings. Eighteen participants replied that it was 'high (making an effort)' while 19 considered it to be 'average (simply completing the task)'. As in study 1, the participants' responses reveal that they did not always read the feedback offered by their peers before recording their podcasts. Twenty-nine students claimed to 'always' read the feedback, seven replied 'sometimes' and only one 'never'.

As for the way in which learners recorded their podcasts, 22 students said that they used a computer while 15 preferred to use their mobile phones. The recording software mentioned included Windows Voice Recorder, Audacity, QuickTime, aTube catcher, and a range of smartphone applications. Most students recorded several versions of their podcast before uploading the final one onto Edmodo. The majority (11 students, 29.7%) said they recorded at least three trials, six students (16.21%) said they recorded four trials and other six claimed to record six trials. Five participants (13.5%) said they usually recorded five, although there were participants who only recorded two trials (3 participants, 8.1%) or even one (2 participants, 5.4%). Finally, there were also individual students (2.7%) who claimed they usually recorded seven, 10, 14 or even 20 versions.

4.2.4 Discussion

The results reported above offer empirical data to provide an informed answer to the research questions addressed in study 1.

RQ1

The first research question (RQ1) addressed podcasts' potential to help learners improve their perception of sounds that are challenging for FL learners: the /s – z/ contrast and English /b d g/ as stops. It was hypothesised that learners receiving training on the target aspects would outperform learners acting as control. Thus, it was hypothesised that G1 would make greater improvements than G2 in their perception of the /s – z/ contrast, and that G2 would progress further than G1 in their perception of English /b d g/ as stops in intervocalic position. The data collected from perception tasks (tasks 1 to 4) offer empirical evidence to support an affirmative answer to this first research question regarding the /s – z/ contrast, but not for /b d g/.

Focusing on the tasks aimed at measuring improvements in the perception of the /s – z/ contrast, the results show that both groups improved their ability to discriminate these sounds from pre- to post-test (task 1). However, only participants in G1 made significant improvements in their ability to identify these sounds correctly (task 2). That is, even though G1 (receiving training on /s – z/) outperformed G2 in these two tasks, the differences between their performances only reached significance in the identification task.

The data from the identification task show that the learners' post-test scores were higher than the ones obtained in the discrimination task (78.8% by G1 and 68.8% by G2). Although the mean scores are still far from the scores obtained by the expert judges (97.5%, 92.5% and 95%), there were some learners who did obtain or even surpass these scores. This indicates that participants were better able to identify instantiations of /s/ and /z/ as belonging to their respective categories than to perceive subtle distinctions between these two sounds in a discrimination task. Nonetheless, when the identification data for /s/ and /z/ were analysed separately, no significant differences were found between sounds. In other words, the improvement made was in the learners' general ability to identify exemplars of the target sounds as either /s/ or /z/, but learners did not show better identification rates for one sound over the other.

A closer inspection of the learners' identification scores for /s/ and /z/ reveals that the mean score for both groups at the beginning of instruction was much better for /s/ than for /z/. The mean pre-test scores for both groups were 14.1 (2.97) for /s/, which represents 70.5% of the total score for this sound, and 10.8 (4.62) for /z/, a 54% of the total score. This confirms the prediction that Spanish learners of English are more capable of identifying instantiations of /s/ correctly, while they have problems in identifying exemplars of /z/, little wonder as the former is a sound they already have in their L1 phonemic repertoire.

A possible explanation for the differences between the discrimination and identification tasks could be found in the influence of orthography or familiarity with the words. Even though participants did not see the written words in this task, they may have resorted to their knowledge about the sound, including typical words and spellings featuring this sound (e.g. <z> as a common spelling for /z/). Nonetheless, participants should not have associated a word like *music* to /z/ simply because they know it is pronounced with /z/. In the instructions they received when performing these tasks, they were asked to concentrate on the sound itself and not on what they knew about the words they were hearing. As for possible influences of spelling, even though some words featured /s/ under the spelling <ce> (e.g. *office, decent*), the majority featured either /s/ or /z/ with the spelling <s>, which is one of the most common spellings for both sounds (e.g. *person, answer, just* for /s/, or *reason, easy, resume* for /z/). The voiced alveolar fricative was only represented by <z> in orthography in two of the novel words (*zero* and *Zoe*), and with the exception of *Xerox* in novel words, the examples where it was featured with <x> can be pronounced either with /s/ or /z/, namely *example* and *exactly* (the two tokens in the testing stimuli were exemplars of /z/).

The mean scores by participants in both groups for the discrimination task were around the 50% performance in pre- and post-tests. This indicates that despite the fact that improvements reached significance, learners were still far from the scores obtained by the three expert judges (95%, 95% and 90%) or the native speakers (95%, 85% and 100%). This suggests that instruction did not help them perceive subtle phonetic differences between these two sounds when they could compare them physically (i.e. it did not improve their perceptual phonetic categories for these two sounds).

As for the improvements in the participants' perception of /b d g/, the results from tasks 3 and 4 do not reveal significant differences for either G1 or G2. The results from task 3 show that even though the experimental group (G2) obtained greater improvements than the control group (G1), whose average scores remained stable or even decreased, the degree of improvement was very modest. The results obtained in the second task (task 4) are in line with those from task 3. Although G1 and G2 scores' are inversely proportional, with G1's average score decreasing 0.6 points and G2 improving in 0.6 points, differences between groups did not reach a level of significance. However, the positive results obtained in the imitation task (task 5) suggest that learners were able to perceive (and imitate) English /b d g/ as stops in intervocalic position. The poor results obtained in tasks 3 and 4 may be due to lack of awareness of their stop realisation in English in task 3, and the learners paying attention to aspects other than occlusion in task 4 (see below).

RQ2

The second research question (RQ2) addressed possible effects of training in the learners' pronunciation of English /z/ (G1) and /b d g/ as stops in intervocalic position (G2). It was hypothesised that if learners in G1 improved their perception of the /s – z/ contrast and participants in G2 improved their perception of /b d g/, they would be able to incorporate these features in their own pronunciation of the target sounds. The results obtained from both groups offer empirical data that support an affirmative answer to this research question, although not for every sound.

Looking at the total scores for each sound across the three production tasks, participants in G1 made substantially larger improvements than those by G2 in their pronunciation of /z/. So did participants in G2 in their pronunciation of /b d g/ as stops. As noted above, both groups showed slight improvements in the aspects for which they were acting as control, but this may be due to the fact that they were receiving training in phonetics at the time of the study. Nonetheless, the results obtained by each group for the target sounds on which they were receiving training were at least twice as positive than those by the control group in the case of /b d g/, and almost three times higher for the group receiving instruction on /z/. Notwithstanding the above, the analysis of both groups' performance in each task reveals that differences in the degree of improvement made between groups only reached significance in the sentence-reading task.

Focusing on the imitation task, both groups made the biggest improvements for the sounds with the lowest scores in the pre-test. G1 learners' pronunciation of /z/ increased by only 0.5 points from pre- to post-test, having started with 48% of the total score. The average scores for G2 on the pronunciation of /z/ increased by 0.8 points, having started with 20% of the total score and showing a slightly greater improvement than the one obtained by the group receiving instruction. In a similar line, the biggest improvement in G2's learners pronunciation of /b d g/ occurred in /b/, the sound with the lowest score in the pre-test (74% as compared to the 90.5% for /d/ and /g/). Even though G2's scores for /b d g/ improved from pre- to post-test, the scores showed ceiling effects from the beginning (74% for /b/, 90.5% for /d/ and 89.5% for /g/). Hence, both groups were able to imitate the stop realisation for /b d g/ in intervocalic position before the beginning of instruction, leaving little room for improvements. Nonetheless, the fact that G1's scores in the imitation task improved so moderately suggests that training did not exert a strong impact on the learners production of /z/.

As for the sentence-reading task, the significant differences found between experimental and control groups reveal a positive effect of instruction. The improvements made by G1 on the pronunciation of /z/ were significantly higher than those by G2. Likewise, the improvements made by G2 on the pronunciation of /b d g/ were generally higher than those by G1. The improvements by participants in G2 on /b/ and /g/ reached significance when considering the scores for familiar items. Nevertheless, when the analysis included novel words for each sound (i.e. the total score for the sound in the sentence-reading task), despite an overall greater improvement by G2, /g/ was the only sound for which improvements reached significance as compared to G1, which is in line with the total production scores.

Finally, the results from the timed picture-description task suggest that training had little effect on the learners' spontaneous production of the target aspects. There were some improvements in the learners' production of these sounds, but both groups obtained similar gains in the pronunciation of both aspects. This suggests that despite the fact learners were able to improve their pronunciation of the target aspects in tasks that were not so demanding (i.e. task 6), they still need more time in order to automatise these improvements.

As explained in section 3.1.3, the difficulty of the above-mentioned production tasks is in line with Morley's (1991, 1994) three modes of practice, namely imitative, controlled and extemporaneous. Good performance in the imitation task should be easier than in the sentence-reading task. This, in turn, should be easier than good performance in the timed picture-description task. The imitation task offers a measure of the learners' perceptual and articulatory abilities, as results from this task only depend on the learners' perception of the sound and their ability to imitate what they hear. However, in the sentence-reading task learners need to be able to produce sounds correctly by recalling them from memory. They need to be aware of the phonological composition of words (often hindered by opaque sound-spelling correspondences) and be able to articulate the target sounds correctly. In this respect, the timed picture-description task imposes similar constraints to the ones faced in a sentence-reading task, although it requires learners to pronounce the target words spontaneously given the cognitively demanding nature of the task and the little time they have to complete it.

The pre-test data for each sound is in line with the hypothesised difficulty imposed by each task. The participants' scores in the pre-test were highest in the imitation task, followed by the sentence-reading task, and finally, the timed picture-description task (see Table 20). However, the same does not hold for the degree of improvement made on each task. Focusing on the improvements made by each group on the aspects they received instruction, the highest improvements in /z/ by G1 occurred in the sentence-reading task, then in the imitation task, and finally, in the timed picture-description task. The biggest improvements by G2 for /b d g/ occurred in the sentence-reading task, followed by the timed picture-description task (except for /g/), and finally, the imitation task. It must be pointed out that this may be due to the ceiling effects in G2's performance on /b d g/ in the imitation task. Thus, there was little room for improvement for G2 in this task. Nonetheless, G1 started with a mean score of 48% in the imitation task in the pre-test but still made smaller improvements than in the sentence-reading task (where they started with 23.3% of the total performance). In any case, the fact that G2 (acting as control for /z/) showed an improvement of 15.2% for /z/ in the imitation task, as compared to the 1.2% or the 2.9% in the other tasks, suggests that it was easier for participants in the control group to make improvements in this task than in the other two.

RQ3

The third research question (RQ3) was concerned with possible generalisation of improvements in words that appeared during training to novel stimuli. It was hypothesised that if learners were able to improve their perception of the target sounds in words that appear in training, they should be able to transfer this improvement to novel words. Generalisation improvements in perception were only assessed for the /s – z/ contrast, as it was assumed that generalisation for /b d g/ should not pose problems given that these sounds are typically represented by the same symbols in orthography (i.e. <b, d, g>). Generalising improvements for /z/ in novel words in production was considered more challenging given that /z/ is represented by various spellings, often coinciding with those for /s/, the sound Spanish learners use instead of /z/. Thus, it may require more time for learners to start realising when /z/ occurs in words they have not seen in training.

The data reported above suggest a positive answer to the question, although only for the /s – z/ contrast. In terms of perception, generalisation gains were observed in the novel stimuli for the /s – z/ contrast in the identification task. However, differences between groups only reached significance for items featuring /s/: 14% by G1 as compared to G2, which remained stable. It is worth pointing out that although learners in the group receiving instruction did outperform participants in G2 in familiar stimuli in the identification task, the improvement made for novel words featuring /z/ was almost the same for both groups (G1 improved 10% and G2 9%). The fact that both groups improved similarly for novel words may be due to the learners' familiarity with the stimuli and the spelling. Familiar items were mostly common words where /z/ is represented as <s> or <x> (e.g. *music, reason, example, exactly, easy*). This may have led learners to identify /z/ as /s/ due to equivalence classification, given that the former does not exist in the phonemic inventory of Spanish and furthermore coincides with spellings that are usually pronounced with /s/ in Spanish (as in *música* /ˈmusika/ or *exacto* /e(k)ˈsakto/). However, novel items were not high-frequency words, probably not so familiar to learners, and participants may have paid more attention to phonetic rather than lexical aspects (e.g. *Xerox, bosom, resume, treason, zero, Zoe, arise, muse*).

The results for the production tasks are rather surprising, as /z/ was the only sound for which improvements in novel stimuli reached a significance level as compared to G2. The improvements made by G2 for /b/ and /g/ were significantly higher than those by G1 in familiar items. Nevertheless, differences between groups were not significant for the scores for /b d g/ in novel stimuli. This contradicts the researcher's predictions regarding /b d g/, as it was hyphothesised that learners should not have problems avoiding to spirantise /b d g/ in novel words once they had mastered these sounds in familiar items, given that they occur in the same positions and are represented by the same graphemes in spelling.

RQ4

The fourth research question addressed the learners' reactions towards using podcasts for pronunciation training. As in the pilot study, the learners' responses to the questionnaires suggest that they enjoyed the project and that they consider this kind of podcast-based instruction potentially advantageous for FL pronunciation training. Overall, they considered the length of study to be appropriate. This time, no student mentioned any of the tasks in the study being too demanding. Since instruction was face-to-face in study 1, all participants completed the tasks they were supposed to complete each week. However, there were still some participants who acknowledged that they did not read the peer feedback received before recording their podcasts. This was the only task that could not be controlled for, as it was done off-campus.

The relationship between perception and production

A comparison between the perception and production data reveals interesting patterns regarding the hypothesised precedence of the former over the latter (see section 1.1.5). For example, the participants' performance in the discrimination task in the pre-test was similar to the one in the imitation task for G1 (48% in task 1 and 48.3% in task 5) and superior in the former for G2 (49% in task 1 and 19% in task 5). Both of these tasks are often used to measure the learners' phonetic categories in phonological acquisition research. However, as noted above, the learners' performance in an imitation task (e.g. task 5) does not reflect differences between perception and production, as the learners' imitation depends on the two domains (see Beddor & Gottfried, 1995). The fact that G1 obtained similar scores in both tasks suggests that these learners' phonetic categories are equally developed in terms of perception and production. Nonetheless, the fact that participants in G2 (acting as control) were more capable of discriminating the /s – z/ contrast (in task 1, with a pre-test score of 49%) than of imitating it (in task 5, with a pre-test score of 19%) offers support to the widely acknowledge claim that perception precedes production (see Flege, 1995). Moreover, the learners' performance in the identification task (task 2) in the pre-test was around 60% of the maximum score, which is also substantially higher than both groups' performance in task 5. In other words, learners were better able to discriminate differences between /s – z/ than to attain /z/ in production when asked to imitate it.

Focusing on /b d g/, both groups' scores were already very high in the imitation task from the pre-test, which suggests that the learners' phonetic categories for English /b d g/ were already adequate from the beginning (with a mean score of 90% for /d/ and /g/ and 70% for /b/). This implies that learners could pronounce /b d g/ as stops in intervocalic position despite the tendency to spirantise them in Spanish. The participants' scores in the imitation task (task 5) in the pre-test were higher than those in the delayed accent-mimicry task (task 3) or those in the identification task requiring them to compare the degree of occlusion in both languages (task 4). This is not surprising for the delayed accent-mimicry task, as learners were required to imitate English-accented Spanish from sentences they read, something very similar to what they had to do in task 6, the sentence-reading task (in fact, the learners' pre-test scores in tasks 3 and 6 are also rather similar). This suggests that even though learners were capable of perceiving and articulating these sounds correctly (as shown in task 5), they did not conceptualise /b d g/ as stops in intervocalic position.

The findings by Rochdi and Mora (2012) suggest that Spanish learners' ability to imitate the pronunciation of English-accented Spanish /b d g/ is related to their capacity to articulate /b d g/ adequately in English, reflecting phonetic-category development. However, the data obtained from task 5 in this study (also commonly used to measure phonetic-category development) are not consistent with the findings from task 3 (whose results were more similar to those in task 6). As Lord (2010) points out, when tackling the contrast between Spanish and English /b d g/, learners do not only need to articulate these sounds correctly in both languages, but also understand the differences between the phonemes and their allophones, as well as their respective distributions.

Notwithstanding the above, the low scores obtained in task 4 are rather surprising, given that if learners are able to produce /b d g/ correctly in an imitation task (which implies an adequate perception of the sound that is imitated), they should be able to perceive occlusion correctly when they hear it. Although both groups' pre-test performance was around 60% of the maximum score for that task (i.e. it was considerably high), the degree of improvement made by the experimental group was expected to have been even higher. It may be the case that learners do perceive differences in occlusion unconsciously when imitating English speech (as evidenced in task 5), but failed to pay attention to the right cues when comparing English voiced stops and their Spanish counterparts. The instruction learners received for task 4 specifically asked them to focus on the degree of occlusion, ignoring aspects such as aspiration, energy of articulation, etc. However, it is possible that learners paid attention to other factors or misunderstood the instructions.

Summing up…

The results show that despite the fact that instruction had a beneficial effect on the learners' pronunciation of the target aspects, with groups acting as experimental usually outperforming the control group, the improvements made could have been much more significant. The differences between control and experimental groups only reached significance for the learners' perception of /s – z/ and for their production of /b/, /g/ and /z/. As noted above, the groups that acted as control for each aspect were also receiving explicit instruction in phonetics at the time of the study, which makes it more difficult for differences to reach statistical significance given that learners in both groups were receiving very specific instruction. However, the average performance by both groups on the target aspects was not near the maximum scores possible for each task, with very little improvement in terms /z/ in the imitation task, or improvements in spontaneous production for either /z/ or /b d g/, for example.

The fact that the two groups in this study acted as both experimental and control offers a double measure of the impact of instruction. However, it should be born in mind that even though participants in the group acting as control did not receive training on the target aspects for the other group, they were enrolled in a phonetics course at the time of the study. This is considered to offer a very reliable measure of the effect of training, given that a control group that received no training whatsoever would be more likely to remain stable as compared to a control group that was simultaneously receiving training in phonetics. Hence, different degrees of improvement between control and experimental groups in this study should be a clear indicator of the impact of the podcast-based instruction.

Limitations and directions for future research

The above suggests that for instruction to yield more positive results, such as fostering automatisation of the target sounds or generalisation to novel contexts, longer instructional periods are needed. It is important to take into consideration that although training in the present study spanned a period of three weeks, it actually took place for an hour a week plus the time learners devoted to recording the podcasts (around 10 minutes). Therefore, training lasted approximately three hours and twenty minutes in total. Other studies training pronunciation explicitly have obtained positive results with as little as four hours of training (e.g. Saito, 2013). However, given that this type of instruction does not offer the high phonetic variability in other types of training (e.g. HVPT), for results to be more positive, longer periods of instruction are probably necessary.

The approach adopted in this study offers learners the opportunity to practise with authentic language in context. Jensen (2011) points out that studies addressing sound contrasts often test the participants' ability to discriminate the contrast but do not offer evidence that training helps them improve their recognition of the target sounds beyond the studies themselves. In this regard, the studies in this followed the opposite trend. Learners were trained with excerpts of real speech and the impact of this training was measured with pre- and post-tests designed to provide an objective measure of the learners' perceptual and productive skills. If learners are trained to perceive and produce the FL phonology with authentic language and are required to produce authentic output in their podcasts, it should be easier for them to generalise what they learn during training to other contexts and situations. However, it is important to note that even if items were considered 'familiar' in pre- and post-test tasks, the number of occurrences of each item during training was sometimes extremely low – with some items occurring only once or twice (see training stimuli section). Thus, even though learners had been exposed to a word during training and could not be considered 'novel', perhaps the frequency of occurrence of the word was not enough for learners so as to gain awareness of the phonological composition of the word.

Instructional periods longer than the ones in the current study would offer the possibility to include a wider range of tasks and opportunities for peer feedback. For example, the combination of scripted and extemporaneous tasks by Ducate and Lomicka (2009). The study by Ducate and Lomicka took place over a period of 16 weeks. This offered learners different types of activities and provided the researchers with a wide range of samples of the learners' speech. Similarly, Lord's (2008) podcasting project was conducted over a semester, with each podcast focusing on a different pronunciation aspect. Participants were asked to create their own podcasting channel and subscribe to their peers' podcasts in order to evaluate them. The original design for the study in this block followed a similar structure (see pilot study). Nevertheless, as pointed out above, there were several limitations that made the researcher devise a much more controlled type of instruction.

Although the contents addressed in the pilot study were closely linked to the oral skills course in which students were enrolled, participants may have not considered it to be part of the course, even if they received marks for it. To a certain extent, this is also the case for study 1. Despite the fact that this assignment substituted one of the regular course assignments, and even though the contents were being covered in the phonetics course students were enrolled, participants were not the researcher's students. This should not have a negative effect on the impact of instruction, but it makes it difficult to implement longer instructional periods as well as to keep control of students do.

There were two aspects that may have conditioned the learners' level of engagement in the pilot study. The first is that they were given the same mark regardless of their performance. This was done in order to obtain a genuine measure of the learners' interest in the project. However, this may also lead some students to simply complete the tasks in order to obtain the reward rather than trying to make the most of the instruction. In fact, this was the impression some of the pre- and post-test recordings in study 1 gave. Some participants completed the pre- and post-test tasks showing little interest in what they were doing. This was aggravated in the podcasts study by the fact that participants in the control group were also given points despite not having to complete training, simply the interviews. As noted above, although this obstacle was overcome in the second study by having both groups act as control and experimental, it may have led some students in the pilot study to simply not make an effort given that they were all going to get the same reward.

The findings from the pilot study show that despite the convenience of offering the whole training online, measuring completion of some tasks was not easy nor suitable for experimental purposes – the biggest issue being that learners did not read the information that was supposed to guide their subsequent listening and recording of the podcasts. Other projects in which learners use podcasts or similar technologies reveal that participation varies widely from week to week, with some students listening to most podcasts but others completing only a few (see e.g. Knight, 2010; Weinberg et al., 2011). While this can be overcome to a certain extent by offering instruction face-to-face (as was the case of the present study), studies delivering materials on campus have also shown that sometimes participants simply do not attend the training sessions (Tanner & Landon, 2009).

For the peer evaluation in the pilot study, participants were given an evaluation sheet with a number of problematic words from the podcasts so that they could concentrate on these when evaluating their peers. Nevertheless, the learners' responses to the questionnaires in the pilot study reveal that they did not feel able to offer valuable feedback to their classmates. In study 1, peer evaluations were simplified by including multiple-choice questionnaires. Nonetheless, the fact that learners were not required to elaborate explicit feedback explaining their peers how to improve limits the opportunities for vicarious learning that arise when learners can see the feedback offered to every participant (see e.g. Swan, 2003). In the pilot study, for example, two participants used YouTube videos to exemplify their peers how to pronounce certain words. Study 1 did not present opportunities for this kind of feedback.

One of the most important things to consider before implementing strategies of this kind is the students' workload (Mason & Rennie, 2008). In this regard, the podcast-based instruction implemented in both studies was considered adequate by most participants, although the approach adopted in the pilot study was much more demanding for both the sparticipants and the researcher. Even though a rubric was created in order to draw the learners' attention to specific features (which allowed them to receive feedback on the same aspects by the researcher and two peers), listening to each podcast and providing detailed feedback on the pronunciation of the whole podcast was extremely time-consuming. Providing detailed individualised feedback for each student could take up to 30 minutes per person, as podcasts were always played several times and specific feedback on how to improve was offered for each learner. This was reduced considerably in study 1, where learners simply had to decide whether they heard the target sound correctly or not in their peers' podcasts but were not required to write feedback on how to correct this. Participants were only told whether their productions of the target sounds sounded right or not. As explained above, articulatory information was given in the explicit instruction offered at the beginning of instruction, or in the first ten minutes of every session if students had questions.

A further limitation of the present study is the fact that learners only had to record two podcasts and that the productions were scripted. This was done in an attempt to alleviate the students' workloads in tasks that were more demanding (as in the pilot study) while prompting numerous instantiations of the target sounds. However, the learners' comments and recommendations to the open-ended questions included the possibility of recording spontaneous podcasts, rather than scripted, or working on podcasts in pairs rather than recording all podcasts individually. As noted above, previous studies using podcasts for pronunciation training have included a mixture of scripted an extemporaneous tasks (Ducate & Lomicka, 2009). Ducate and Lomicka did not find substantial differences in comprehensibility or accentedness between the extemporaneous and scripted tasks. Nonetheless, learners in the present study obtained considerably lower scores in the spontaneous task (task 7) than in the controlled task (task 6), which suggests that including a mixture of both types of activity may foster greater improvements. Future implementations podcasts for pronunciation training should include a higher number of speaking tasks and of a wider variety, like the ones adopted by Lord (2008) or Ducate and Lomicka (2009), for example. Given the benefits tandem exchanges offer for pronunciation practice (Horgues & Scheuer, 2015), an interesting possibility would be to create podcasting communities with international students where FL learners could record podcasts for native listeners and viceversa.

Another limitation of the current study was the impossibility to make comparisons from the learners' responses at the pre- and post-test questionnaires. As noted above, learners were not asked to write their participant number in order to obtain a genuine measure of their perceptions towards the podcast-based instruction. If participation is anonymous, learners should not feel pressured to save face with the researcher or the course lecturers. Nevertheless, it would have been interesting to investigate possible changes in their perceptions towards their pronunciation abilities (e.g. perception, production, giving feedback), shyness or self-confidence from pre- to post-test.

In addition, even though participants were rewarded based on task completion, the tools employed for activities 1 and 2 and for the peer-feedback (Edmodo and Google Drive questionnaires) did not allow the researcher to easily measure their engagement qualitatively. Participants were given the answers to activities 1 and 2 after they completed the activities, but it was not possible to observe whether learners were really making an effort or they were simply filling in the questionnaires. Some participants failed to identify supposedly clear exemplars of the target sounds in some of the activities, which suggests that perhaps some students completed some of the tasks simply to be rewarded. Approaches that require learners to write constructive feedback themselves, like the one used by Lord (2008) or the one adopted in the pilot study, would allow teachers and researchers to base their evaluations on the quality of the feedback provided. For future implementations of studies of this kind, it would be interesting to measure the learners' capacity to provide feedback to their peers, as this would make them reflect on language, check their understanding of the target aspects (see Couper, 2011), and eventually, it would hopefully help them self-monitor their own progress.

Finally, since the ultimate goal of instruction with podcasts would be that learners become capable of noticing features in the input they are exposed to and try to implement them into their speech, future approaches could attempt to offer explicit instruction at an initial stage and then encourage learners to pay attention to these features by simply listening to the podcasts, without additional activities (like activities 1 and 2 in study 1). If learners start noticing how people pronounce the features under investigation by simply listening to podcasts, it should be easy for them to transfer the benefits obtained from instruction to other situations, such as listening to native speakers talk, watching films or TV, etc. Moreover, this would enable learners to listen to podcasts anywhere. So far, the majority of participants in studies of this kind tend to listen to podcasts at home, therefore not exploiting the portability for which these tools were conceived (see e.g. Knight, 2010). It would be interesting to explore whether podcasts can promote perceptual gains when listening to them 'on the go'.

Chapter V: Exploring smartphone apps' potential for autonomous pronunciation training

This chapter addresses the potential of smartphone apps for autonomous pronunciation training. In particular, the chapter weighs the possibilities this technology offers in order to improve FL learners' perception and production of a number of sounds that are problematic for EFL learners. In an attempt to drift away from the common laboratory-like settings in which research is often conducted, this study investigated the potential of a commercial smartphone app when learners use it outside classroom settings.

5.1 Study 2

5.1.1 Aims, research questions and hypotheses

Given the impossibility to test an app developed by the researcher, study 2 in this dissertation explores the potential of the *EFP* app (OUP, 2012) to improve Spanish learners' pronunciation of a number of problematic English phonemes. In particular, it addreses three of the specific objectives in this dissertation, namely to assess the potential of smartphone apps to improve:

4. The potential of smartphone apps to improve Spanish learners' perception of English /æ/, /ʌ/, /ɑː/, /ə/, and the /s – z/ contrast.
5. The potential of smartphone apps to improve Spanish learners' production of English /æ/, /ʌ/, /ɑː/, /ə/, and /z/.
6. The learners' perceptions towards using smartphone apps for pronunciation training.

In order to meet these objectives, study 2 addresses six research questions and five hypotheses. They are based on the advantages of smartphone apps described in section

1.1.3 and on the importance of focus on form and the advantages of teaching phonemic symbols stated in section 1.1.5.

RQ1: Can instruction with an app foster improvements in the learners' perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/?

H1. Training with the app will help learners improve their perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/.

RQ2: Does perceptual training with an app foster improvements in the learners' production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/?

H2. Perceptual training with the app will help learners improve their production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/.

RQ3: Does training with the app generalise to novel contexts?

H3. Training with the app will help learners improve their perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/ in novel words.

H4. Training with the app will help learners improve their production of /æ/, /ʌ/, /ɑː/, /ə/ and /z/ in novel words.

RQ4: Does training have the same impact on learners enrolled on an English Studies degree taking a phonetics course than on learners from other degrees not receiving training in phonetics?

H5. Training with the app will foster similar improvements in English majors taking a phonetics course than on learners not receiving training in phonetics.

RQ5: What are the students' reactions towards using apps for pronunciation training?

5.2.2 Method

a) Participants

Sixty-three participants took part in this study. They were 15 male and 48 female (mean age 19.2; SD = 0.88).[91] As the other studies in this dissertation, this study was advertised as a free pronunciation course. This time, it was advertised to students enrolled on different English courses from various degrees at the University of Murcia and to the same group of

[91] The data about their mean age has been obtained from the 59 respondents who completed the initial questionnaire. Despite the researcher's insistence on the importance of completing the questionnaire, there were four students who did not do it.

students that participated in study 1 (enrolled on a phonetics course in the second year of an English Studies degree). Students in the English courses were first year students from the degrees in Translation (English major), Hispanic Studies, and Medicine.

The recruitment process took place during the first (autumn) semester of the academic year 2014/2015. This allowed conducting the study at a time students were not too stressed due to the exam period, which may yield a high number of drop-outs, as happened in the podcasts pilot study. This study was initially planned for the same group of participants in study 1, as it could be included as part of their phonetics course given the relevance of the materials for the course. As noted in section 4.2.2a (study 1), some of these students had already participated in the podcasts pilot study as part of an oral skills course in the first year of their degree. An effort was made to allocate participants who acted as control group the pilot study to study 1, as they had been offered this opportunity the previous year. Those who had already acted as experimental group, and had therefore received training through podcasts, were assigned to study 2 in order to offer them a different type of training.

As in study 1, participants from the English phonetics course were awarded 0.7 points of the final mark if they completed all the tasks, as participation in this project replaced one of the assignments of their course (see section 4.2.2a for details). Participants were informed that they would be evaluated based on task-completion, not on the correctness of their responses in the different activities. Those students who volunteered to participate from the Translation, Hispanics and Medicine degrees were not awarded any points for their participation, given that despite the importance of pronunciation in any English course, such specific pronunciation instruction was not part of their syllabus. Nevertheless, as the rest of participants in this study, they were given the app for free (see details below).

The idea of including students from different degrees was to test the potential of this technology with a group of students who were not receiving instruction in phonetics at the time of the study. Students from the English Studies degree should already be familiar with phonemic symbols, a central element in this study, and may therefore have some kind of advantage over students from other degrees. However, since the main aim of this dissertation is to offer alternative ways to train FL pronunciation to the average language learner, this small group of volunteers from other degrees served to test the potential of training with people who were not studying phonetics. It should be noted that students from the degree in Translation finish their studies with a C1 level according to the CEFR, the same level as those in the English Studies degree. Nonetheless, the degree does not include specific training in phonetics and, at the time of the study, they had just begun their degree.

Participants were divided into three groups. Two groups acted as experimental groups (groups 1 and 3) and one acted as control (group 2). However, after the experimental groups had finished training, group 2 started receiving instruction, therefore also serving as experimental.

There were 52 students from the English Studies degree and 11 students from the other degrees (seven from the Translation Degree, three from Hispanic Studies and one from Medicine). Participants from the English Studies degree were divided into experimental (group 1; n = 25) and control (group 2; n = 27) groups. Participants from the other degrees (group 3; n = 11) also acted as experimental at the same time as group 1, but they will be referred to as group 3 in order to differentiate them from the different type of profile of participants in groups 1 and 2.

b) Training procedure

As mentioned above, training consisted in using the *EFP* app (OUP, 2012), henceforth EFP, over a period of two weeks following the instructions provided by the researcher. The app cost 5.5€ at the time the study was conducted, but participants were informed that they would be given the app for free in exchange for their participation in the study. The app was reimbursed to those participants who completed all tasks after the last post-test for their group. In order to acknowledge reimbursement of the app, participants were required to sign an acknowledgment of payment sheet indicating the amount reimbursed and the date (see Appendix 13).

The EFP app includes a sound chart with the sounds of English (British and American) and two activities that provide practice with those sounds. The sound chart employs the same phonemic symbols with pictorial illustrations offered in the *New English File* collection of books (e.g. Oxenden, Latham-Koenig, & Seligson, 2004) – see Figure 50. When users tap on a symbol, the sound is played in isolation and in a sample word. If the symbol is tapped on a second time, users are directed to a different screen where they can hear the sound in context. They can choose to hear it in five words (Figure 50, middle) and five sentences that show the spelling for the sound in a different colour (Figure 50, right). Additionally, users are also offered the opportunity to record themselves and compare their recording to the model stored – one of the traditional functions commonly offered in pronunciation courseware (see chapter I).

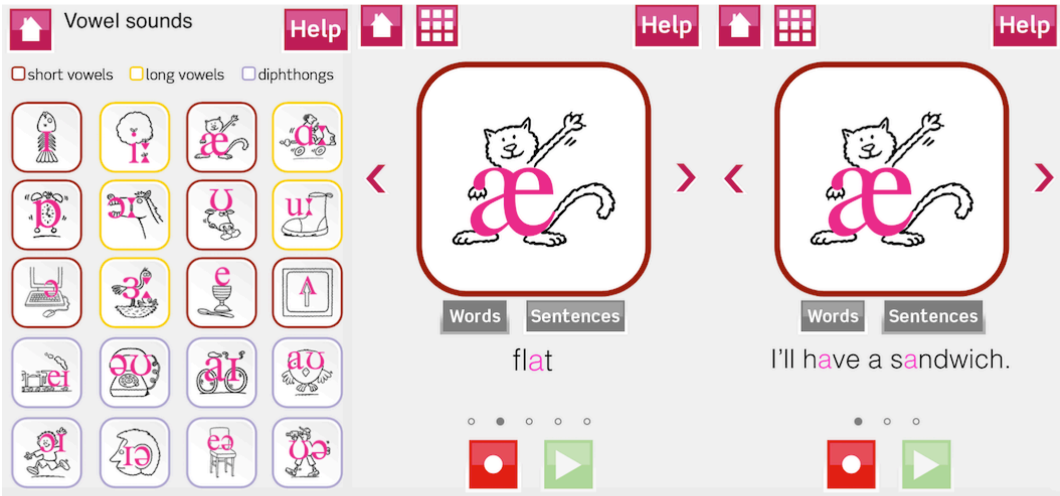


Figure 50. Screenshots from the EFP app: Sound chart (left), sample word (middle) and sample sentence (right)

The first activity in the app is a sound identification activity. In this activity, words are played one by one and users have to identify which of two possible phonemes (displayed as phonemic symbols) is the one they are hearing in that word. The orthographic representation of words is not shown (Figure 51, left). Users are offered immediate feedback on the choices they make. They are shown a green tick if their answer is correct, a red cross if their answer is wrong (Figure 51, middle). Every 10 words, a progress screen shows the score for those 10 words as well as a summary of the user's responses. This summary shows the learners' scores on each word, including their phonemic transcription, therefore also providing users with the right answer (Figure 51, right).

In the second activity, words from activity 1 are presented individually and users need to choose between two possible sounds for each word. This time, users are shown the spelling of the word, but they cannot hear the word. They need to recall what they have learnt in the previous activity (Figure 52).

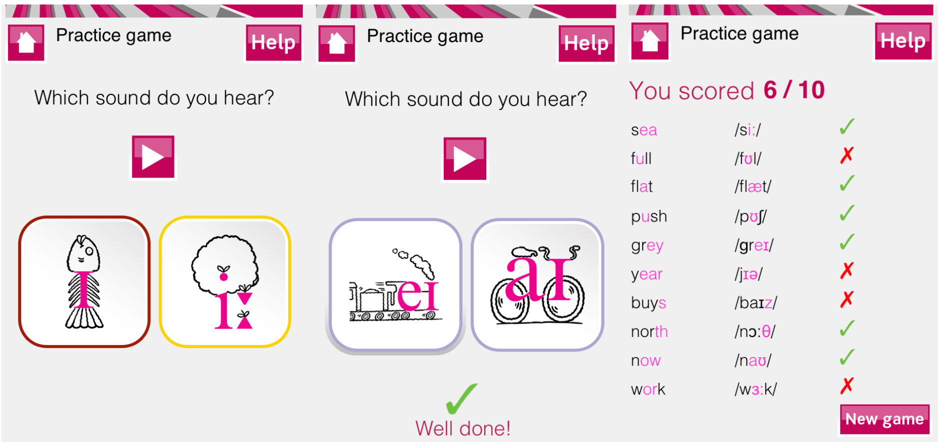


Figure 51. Screenshots from the EFP app. Activity 1 (left and middle) and progress screen (right)

In line with the other studies in this dissertation, this study intended to depart from the traditional laboratory-like settings in which CAPT research is often conducted. Nevertheless, the way learners used the app was not completely autonomous, as they were given guidelines on how to use it and their participation was being controlled for. Some of the studies exploring the potential of apps have not given students guidelines about how to use the app (Amer, 2014), or have offer them the possibility of using either the app or the computer version of it (Li & Hegelheimer, 2013; Stockwell, 2007, 2010). In this study, some kind of control over learners' performance was considered necessary in order to make any claims on the potential effects of training – as learners' usage of the app should be as similar as possible.

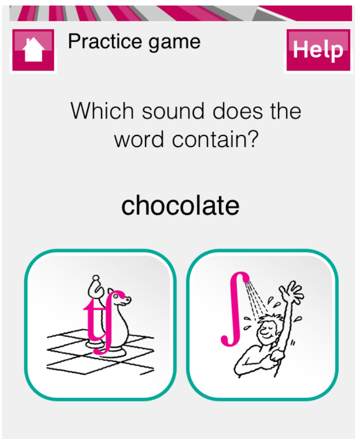


Figure 52. Screenshot from the EFP app: Activity 2

A pilot test of the app was carried out with with three volunteer students over the summer. These three students were given a free download of the app in exchange for their collaboration. They were required to complete the pre- and post-tests in order to measure the amount of time required for each test as well as to use the app over ten days in order to test the procedure. The researcher met with the participants for the post-test and had an informal interview with them. Their suggestions were not to encourage use of the app for more than 15/20 minutes a day, as some of the words tended to be repeated numerous times and the activities could be perceived as monotonous.

Hence, at the beginning of study 2 participants were given a brief induction on how they should use the app after the group interview in the pre-test. They were instructed to use the app for around 15/20 minutes a day, five days a week (from Monday to Friday), over a period of two weeks. Thus, learners received instruction over a total of 10 days. Nevertheless, controlling the time spent on task would not be a reliable measure for homogeneity across participants, as the time spent completing the activities will never be the same for every learner. Therefore, they were told that they had to complete 10 games on each of the two activities every day (see Figure 51, right). Each game included 10 stimuli. Thus, they were exposed to 100 audio instantiations on a range of English sounds a day in activity 1, and they were tested their knowledge of the phonological make-up of 100 words a day in activity 2. For the sake of simplicity, every ten games (i.e. 100 words) have been considered 'a level'.[92]

In the induction session participants were told what the target sounds in this 'free pronunciation course' were going to be. They were instructed to explore these sounds in the phonemic chart before every day's training, as some sort of warm-up activity. They were encouraged to tap on the symbols and compare the different target sounds in order to try to become aware of their differences, and also to explore the sounds in the sample words and sentences. This would allow learners to be exposed to the sounds in different positions, hear instantiations of the sounds in words and sentences, and also see some common spellings for those sounds. Additionally, as in the study by Saran et al. (2009), participants were strongly recommended to imitate the pronunciation of the words in the phonemic chart and activity 1 out loud in an attempt to foster production practice (uncontrolled). Finally, they were told to always complete activity 1 first, as activity 2 was a revision of the words previously presented in activity 1.

Edmodo was also used in this study to keep in contact with participants during the two weeks of training. Learners were not required to interact with the researcher during the study, but they were given the opportunity to use this tool in case they had any questions

---

[92] It should be noted that unlike the usual pattern in games, progression through the different 'levels' here does not imply increasing difficulty. This is just a classification made by the researcher in order to quantify the learners' participation.

or comments. The wall in Edmodo is similar to a forum, so it was considered a handy tool in order to solve doubts that may be of interest for the whole group and therefore act as some sort of common ground for discussion with potential for vicarious learning.[93] The PowerPoint explaining the weekly procedure used in the induction session was also uploaded onto Edmodo as a .pdf file so that students could access it any time during training (see Appendix 14).

In order to control task completion, participants were asked to take a screenshot of every progress screen they reached each day. At the end of the day, participants needed to have a total of 20 screenshots, 10 from activity one and 10 from activity two. This method allowed checking the times at which activities were completed, as well as the dates of completion. Progress screens in this app (Figure 51, right) disappear after users start a new activity and they cannot be retrieved again. Thus, since screenshots show the time and date at which they were taken, this ensured that learners completed every activity on the day they were supposed to. Unfortunately, there was no way of checking completion of the warm-up activity.

In order to monitor the learners' participation, these were required to create a Dropbox account (if they did not have one) and download the Dropbox app onto their smartphones.[94] Instructions on how to do so were given during the induction session after the group interview in the pre-test. Users were told to create a folder with their surname and name in Dropbox and share it with the researcher. They were advised to create an additional folder for each day, as this would allow an easier classification of screenshots (Figures 53 and 54), although this was optional given that screenshots already showed the date and time. This method allowed them to upload pictures onto a folder the researcher would be able to access any time, having an up-to-date measure of their progress.

[93] The type of learning that occurs in classroom when students learn from clarification other students' doubts also takes place in online interactions. This is not restricted to when learners receive posts from the teacher directly, but also by being observers of other interactions and learning from them (Swan, 2003).

[94] Dropbox is an online storage service that allows users to save data for free. It has a desktop version and an app version, but the app version was considered more convenient as it allowed learners to upload their daily screenshots directly from their phone or tablet – the device where they were using the app.

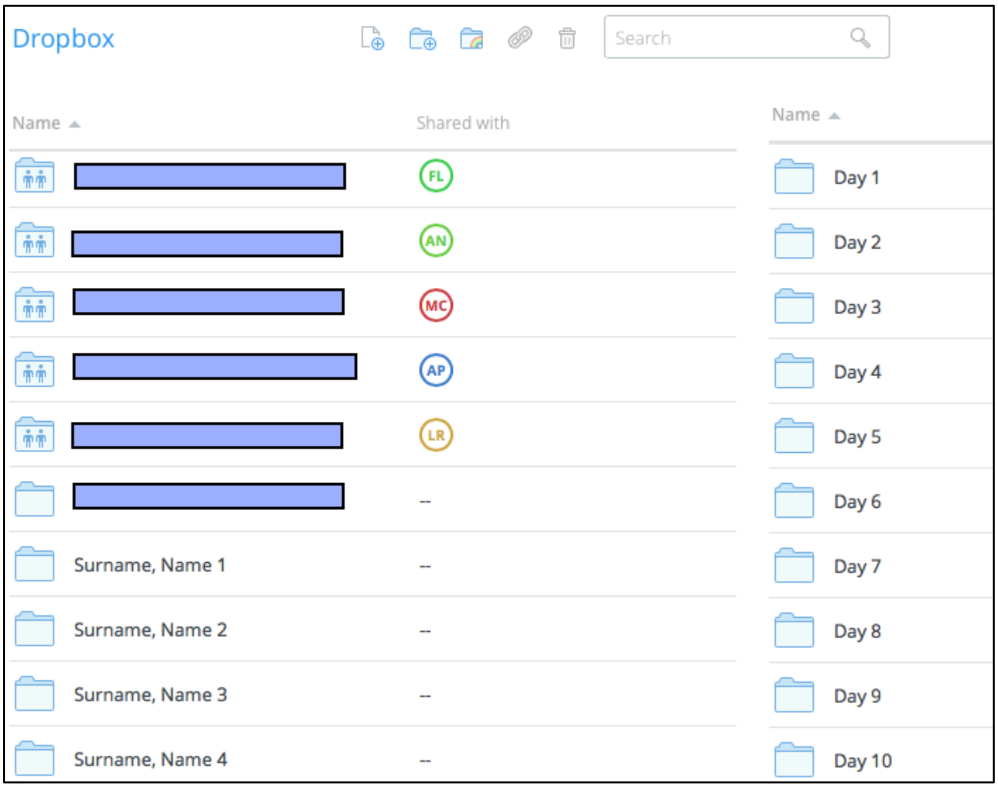


Figure 53. Arrangement of shared folders in Dropbox (study 2)

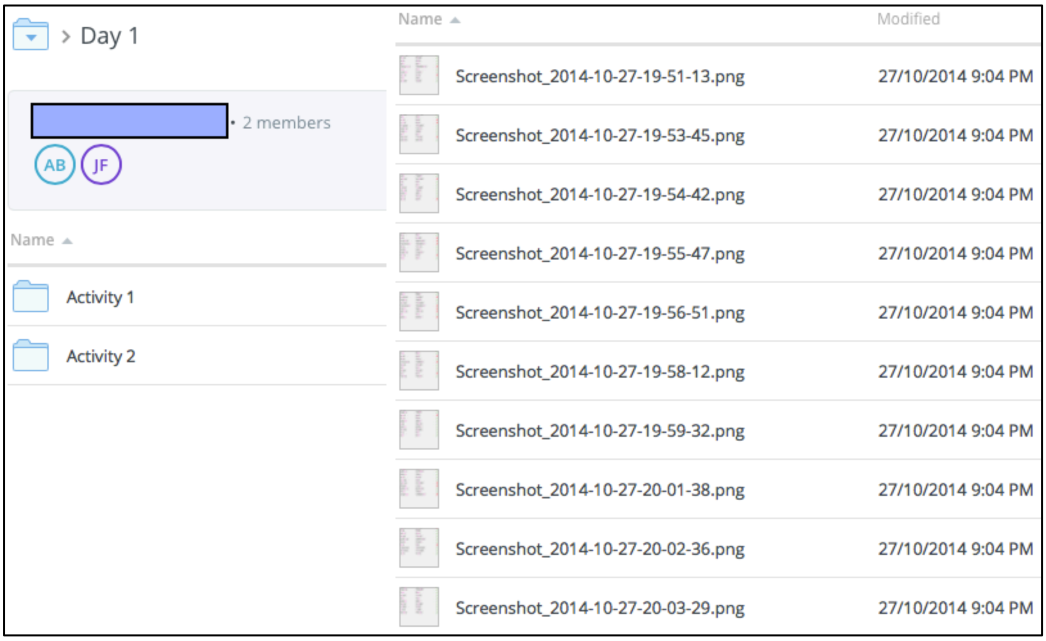


Figure 54. Sample of screenshot classification in Dropbox (study 2)

c) Training stimuli

As in study 1, it must be pointed out that training stimuli could not be selected by the researcher. Stimuli in pronunciation acquisition studies are often very controlled in terms of phonetic environment and number of appearances of each stimulus. However, stimuli in this study were limited to the input offered by the EFP app. One of the limitations of using materials that are not designed by the researchers themselves is that the input to which students are exposed cannot be controlled for. Moreover, the EFP app does not allow learners to select target items with which they want to practise, as opposed to apps like *Sounds* (Macmillan Publishers Ltd., 2011), for example. In the EFP app, learners are presented with randomised examples of words featuring a range of different English phonemes.

In light of the above, the researcher tested the first 1000 stimuli in the app for activity 1 (with sound) and the first 500 for activity 2 (without sound). This was done in order to calculate the approximate amount of input learners could be exposed to when using the app. A record was made of those 1500 tokens in the two activities, calculating the percentage of appearances for tokens featuring each sound. It is important to note that this can only be taken as a tentative observation, as it is not possible to predict all the possible appearances of each word in the app given that the process of stimuli presentation is unknown to the researcher. Words do not always appear in the same order, and even if they did, this could change if the app were updated. However, it became apparent during the testing procedure that patterns of occurrence of sounds are fairly regular. Even though words featuring each sound may not always be the same, or even appear in the same order, it is believed that the calculated rate of exposure users may have of each sound is reliable.

It must be noted that the EFP app is devoted to familiarising users with the sound system of English. Consequently, the target sounds in this study are only part of all the sounds addressed. Hence, of the 10 words that make up a game, only a few exemplify the target sounds in this study. Figure 55 shows the percentage of appearance of each sound after 1000 stimuli for activity 1 and after 500 stimuli for activity 2.[95] As can be seen, the frequency for target sounds is fairly balanced in relation to other sounds, with the exception of schwa, which is considerably higher. Figure 55 includes only those sounds that are minimally represented in activities 1 and 2. Even though the app includes all the English sounds in the phonemic chart, sounds that do not appear in the figure are

[95] Since activity 2 is a revision of the words appearing in activity 1, it was assumed that the representation of each sound in the two activities should be similar. In fact, the data on Appendix 15 show that percentages of appearance in activities 1 and 2 (obtained from 1000 and 500 stimuli respectively) do not differ in more than 1.5 points (see Figure 56 below). Thus, despite having been calculated over a smaller sample, the data obtained from activity 2 seems to be in accordance with the data obtained for activity 1.

completely neglected in the activities (see Appendix 15 for a list showing the percentages of appearance for all sounds).

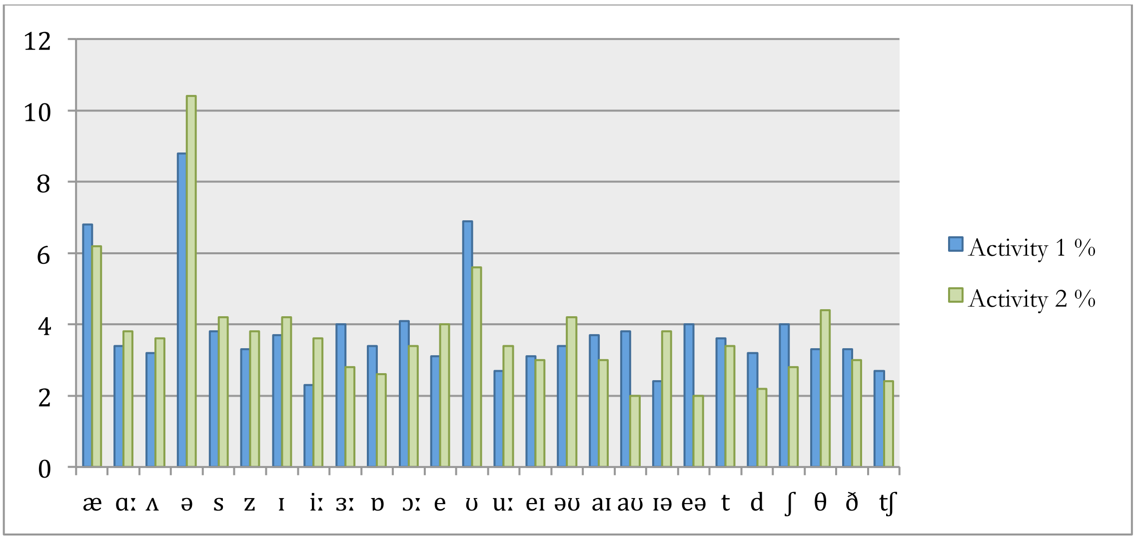


Figure 55. Percentage of occurrence of each sound in activity 1 (out of 1000 stimuli) and activity 2 (out of 500 stimuli)

It is important to point out that, since participants completed only 10 games every day on each activity, the representation of each of the target sounds in those 10 games was rather small. Table 23 illustrates the percentage of appearance of each sound every 100 words (the daily exposure learners received on each activity). The two most represented sounds were /ə/, appearing 88 times in activity 1 and 52 in activity 2; and /æ/, featured in 68 words in activity 1, and in 31 in activity 2.

Table 23 presents the number of instantiations and the percentage of occurrence of each sound over a set of 1000 (10 levels) and 500 (5 levels) stimuli for activities 1 and 2 respectively. The 10 levels the researcher completed in order to analyse the stimuli correspond to the 10 levels learners would complete with the whole training.

Table 23. Number of instantiations and percentage of occurrence of the target sounds in the EFP app

	Activity 1		**Activity 2**	
	(*n* = 1000)	**%**	**(*n* = 500)**	**%**
/æ/	68	6.8	31	6.2
/ɑː/	34	3.4	19	3.8
/ʌ/	32	3.2	18	3.6
/ə/	88	8.8	52	10.4
/s/	38	3.8	21	4.2
/z/	33	3.3	19	3.8

As regards frequency of occurrence of one sound in relation to the others, the number of occurrences of target sounds in the app was rather balanced, with the exception of /ə/and /æ/ that were slightly higher (see Figure 56).

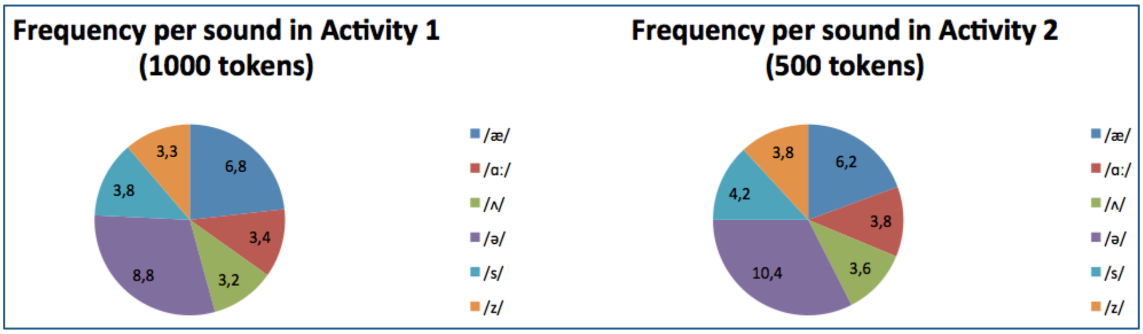


Figure 56. Frequency of occurrence of each target sound in activities 1 and 2

However, a closer inspection of the number of instantiations of each sound per level reveals that, despite general trends confirming the figures above, some levels (the amount of training learners would receive daily) include a higher number of exemplars for each sound than others. Figures 57 and 58 show the number of words featuring each sound that appeared every level in activity 1 (Figure 57) and in activity 2 (Figure 58). The figures have been obtained using the 10 levels (1000 stimuli) analysed by the researcher for activity 1 and the 5 levels (500 stimuli) for activity 2 as a reference. The specific figures showing the number of tokens featuring each sound per level can be found in Appendix 16.

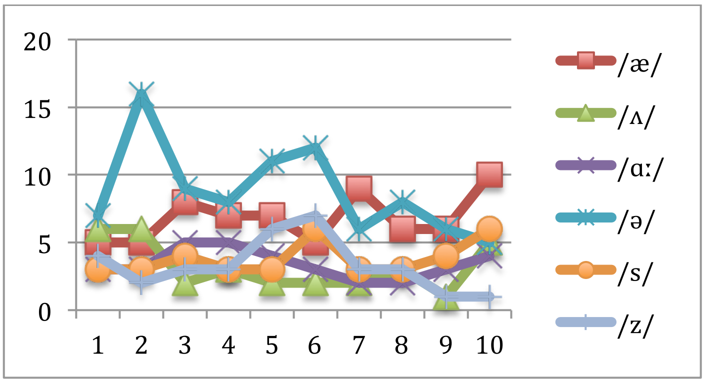


Figure 57. Number of occurrences of each sound (vertical axis) per level (horizontal axis) in activity 1

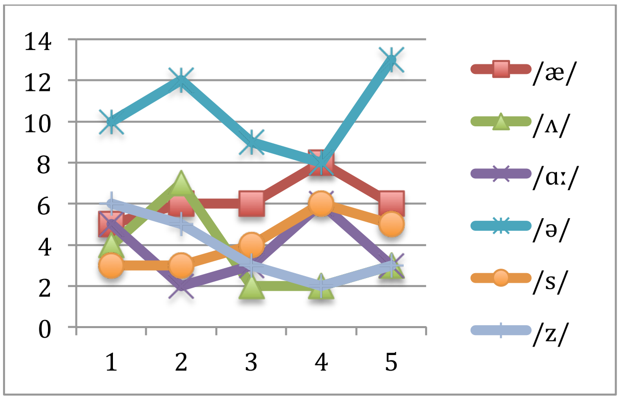


Figure 58. Number of occurrences of each sound (vertical axis) per level (horizontal axis) in activity 2

Words featured the target sounds spelt in different ways and in different positions (initial, medial and final – though not for every sound). Appendix 17 shows the list of words used, including the number of repetitions of each word in these samples. They are arranged in order of frequency of appearance, as this was one of the main criteria for the inclusion of these items as part of the testing stimuli (see below).

d) Testing procedure

Participants in this study had to complete a total of four interviews, two at the beginning and two at the end. However, participants in group 2, who acted as control and experimental group, had to meet the researcher for an extra interview. In order to avoid imposing excessive demands on this group as compared to the other two, they were only required to complete the perception post-test. Asking them to complete an additional production test was considered excessive given that their participation in the study had already been two weeks longer than that of their classmates, even without having received training during the first two weeks. Additionally, the production sample obtained from the experimental groups 1 (n = 21) and 3 (n = 11) was considered adequate in order to provide an informed answer to RQ2, as other studies have offered valuable insights on production improvements with similar or even smaller numbers of participants in their experimental groups.[96]

The pre-test was administered to all groups at the beginning of the study. Groups 1 and 3 received training during the first two weeks of the study, after which all groups completed a post-test. After group 2 had served as control, participants in this group started training. Training for group 2 also lasted two weeks, after which they completed a final post-test, this time only addressing perception. This arrangement is illustrated visually in Figure 59.

1. Pre-test (Groups 1, 2 and 3)
    a. Perception
        i. Discrimination task
        ii. Identification task
    b. Production
        i. Sentence-reading task
        ii. Picture-description task
        iii. Imitation task
2. Instruction (Groups 1 and 3)
3. Post-test (Groups 1, 2 and 3)
4. Instruction (Group 2)
5. Second post-test (Group 2)
    a. Perception
        i. Discrimination task
        ii. Identification task

Figure 59. Outline of the procedure followed in study 2

The perception test included a discrimination and an identification task. The production test consisted of three tasks, an imitation task, a sentence-reading task and a timed picture-description task (sections 3.1.2 and 3.1.3).

For the production tests, participants met individually with the researcher for around 10 minutes (10 mins x 63 students x 2 testing times = approximately 21 hours). Perceptual tests were administered in groups of around 17 participants in the same computer room described above. Meetings for the perceptual test lasted about an hour, as learners were also given a brief induction explaining how they had to proceed during the two weeks of training.

[96] Lambacher et al. (2005) used 20; Lord (2008) used 16 and Lord (2010) used eight; Saito (2013) used 17 in his FFI + EI group and 18 in his FFI only group.

As in study 1, before starting the discrimination test, participants were trained with an oddity discrimination task using geometric shapes and were explained the different parts of the test with examples containing non-target sounds (see section 4.2.2e).

The identification task followed the same procedure as that in study 1. As explained in section 3.1.2, labels for the /s – z/ contrast were not considered problematic since the phonemic symbols and the most common spelling of those sounds coincide. However, labels for vowel sounds were accompanied by a keyword that consisted of a high-frequency sample word featuring that sound (e.g. /ɑː/ as in *bar*; /ʌ/ as in *up*, and /æ/ as in *cat*).

The voices for the stimuli presented in the pre- and post-tests were always different from the ones in the input learners received during training. Furthemore, testing stimuli included familiar and novel words to test possible generalisation gains to words that did not appear in training.

The identification test was divided into three activities in TP (see Figure 60 below). The first and second addressed vowel sounds in different positions and the third focused on the two target consonants.

Activity 1 in the identification task assessed the learners' identification of the target vowels in words featuring the sounds in different contexts and under different spellings. Hence, it followed the same pattern as activity 1 in the app. Learners heard a word and they had to decide which of the vowels presented they were hearing. Students were not shown the spelling for the words in this activity, only the auditory input. This guaranteed that they were actually 'identifying' the sound, rather than associating a common spelling with it (although this possibility cannot be discarded if participants recognise the word and recall which sound must be associated with a particular spelling – e.g. /ə/ for the <er> in *dinner*). Even though the task occasionally included words with several syllables (and therefore different vowels), the non-target vowel was never one of the choices in the identification task. For example, learners would hear the word *dinner* with /æ ʌ ɑː ə/ as options, but never including /ɪ/.

The options for identification were kept invariable in the tasks. There was one set of options for vowel sounds (/æ ʌ ɑː ə e/ + 'I don't know') and another set for consonants (/s z ʃ/ + 'I don't know').

The second activity was meant to assess the learners' ability to identify the target sounds in contexts including vowels that were phonetically similar. They were featured in a wider range of contexts and under different orthographic representations. Tokens contained several target vowels, therefore often including two of the possible responses provided too (e.g. *another* – /ʌ/ vs. /ə/). Hence, for this activity participants were given a list of 27 words with the target vowels underlined and in bold (Appendix 18). Learners had to listen to each word individually on TP and identify which of the given options corresponded to the underlined vowel in that word. It should be noted that the words *glasses* and *present* were considered as 'familiar' stimuli because they appeared in the app. However, in the trial run with the 1000 stimuli explained above, these two words did not appear in the activity with audio; they only appeared once and it was in activity 2. Hence, even though these items were considered 'familiar', learners did not receive aural exposure to them during training.

Finally, the last task addressed the learners' perception of the /s – z/ contrast. In the familiar stimuli included in the app, the contrast is limited to word-final position, except for *museum, music* and *exams*. Thus, tokens featuring familiar stimuli were also reduced to those contexts, although novel stimuli included the contrast in initial (n = 2), medial (n = 4) and final (n = 4) position. As in study 1, in this final task for /s – z/, six distractors were also included featuring the voiceless palato-alveolar fricative in every position, with two tokens for each position.

e) Testing stimuli

Perception

Stimuli for the perception tasks were obtained from the dictionaries mentioned in section 3.1.1, with the exception of 13 items in the identification task featuring the /s – z/ contrast in plural words. These words had to be included in the test given that they were among the 'familiar' stimuli from training (see Appendix 17). Since audio illustrations in pronunciation dictionaries do not include plurals, these words were recorded by a female native speaker from Brighton and a male native speaker from Preston (UK). These audios were edited following the same procedure as for the rest of audios in terms of normalisation and reduction of background noise.

Stimuli for the discrimination task consisted of 40 triads of minimally paired words and 10 triads with the strong and weak versions of words whose vowels can be reduced to schwa. There were 10 minimal pairs for the /æ – ʌ/ distinction, 10 for /æ – ɑː/, 10 for /ʌ – ɑː/, 10 for /s – z/, and 10 triads with the weak and strong versions of words containing schwa in their reduced form (see Table 24). Some studies have limited vowel distributions to a small number of contexts in order to reduce coarticulation effects on the target vowels (e.g. Baker & Trofimovich, 2006). Nevertheless, given the variety of contexts in which target sounds appear in the EFP app, it was considered important to include examples of sounds in a variety of phonetic environments in order to test the learners' perceptual categories for the target sounds regardless of the context.

Minimal pairs for vowel sounds were monosyllabic words with the target vowel as nucleus in word-medial position. Tokens included nuclei surrounded by two voiceless consonants (e.g. *hat-hut; cap-carp*), by two voiced consonants (e.g. *bag-bug; bun-barn*), by a voiced and a voiceless consonant (e.g. *bat-but; hat-heart*) or vice versa (e.g. *come-calm; cad-card*). In initial position, vowels were also followed by voiced or voiceless (e.g. *app-up; am-arm*) consonants. As regards /ə/, it was difficult to obtain minimal pairs only differing in this sound, except for those words that have a strong and a weak version, where the full vowel in the strong version is reduced to /ə/ in unstressed positions. Thus, triads for schwa included the same word three times, comparing weak and strong versions of those words (e.g. *and* /ænd/ vs. /ənd/; *that* /ðæt/ vs. /ðət/). For the /s – z/ contrast, words were selected following the same criteria as in study 1, featuring this contrast in initial, medial and final position.

Table 24. Stimuli for the discrimination task in study 2.

	/æ/	/ʌ/		/æ/	/ɑː/		/ʌ/	/ɑː/		/ə/		/s/	/z/
1	**hat**	hut	**11**	**cap**	carp	**21**	**cup**	carp	**31**	**and**	**41**	**sap**	zap
2	**bag**	bug	**12**	hat	**heart**	**22**	**hut**	heart	**32**	**but**	**42**	seal	**zeal**
3	**bat**	but	**13**	**ban**	barn	**23**	bun	**barn**	**33**	**that**	**43**	sing	**zing**
4	matt	**mutt**	**14**	pack	**park**	**24**	**come**	calm	**34**	**at**	**44**	said	**z**
5	cap	**cup**	**15**	**pat**	part	**25**	buck	**bark**	**35**	**have**	**45**	**racing**	raising
6	track	**truck**	**16**	cad	**card**	**26**	**huff**	half	**36**	**must**	**46**	fussy	**fuzzy**
7	**app**	up	**17**	**hack**	hark	**27**	pus	**pass**	**37**	**than**	**47**	**muscle**	muzzle
8	**ban**	bun	**18**	am	**arm**	**28**	**done**	darn	**38**	**some**	**48**	**precedent**	president
9	bank	**bunk**	**19**	**bat**	Bart	**29**	cluck	**clerk**	**39**	**can**	**49**	**bus**	buzz
10	pat	**putt**	**20**	chat	**chart**	**30**	dunce	**dance**	**40**	**does**	**50**	price	**prize**
				Distractors /s - ʃ/									
51	**sue**	shoe	**53**	seesaw	**seashore**	**54**	**Iris**	Irish					
52	see	**she**				**55**	mass	**mash**					

A total of 80 triads were created for the discrimination test. There were 50 change triads (e.g. *cap-cup-cap*), including 10 for each pair of target sounds and for schwa, and 25 catch triads (e.g. *cap-cap-cap*), with 5 for each target sound except for schwa. Schwa could not be featured in five catch triads due to the impossibility of obtaining three different weak versions of words with schwa in the dictionaries mentioned above (e.g. three instantiations of the weak version of *that*). Five more triads were included as distractors, featuring the /ʃ – s/ contrast. As in study 1, each option '1, 2, 3' appeared as the correct option the same number of times (18) to avoid response biases. The option for catch triads (i.e. 'the same') appeared slightly more often (25 times) in order to include at least 5 catch triads for each individual sound (except for schwa).

Every sound appeared in the 'odd' position the same number of times. However, schwa always had to appear in the odd position due to the lack of adequate available stimuli featuring the weak versions of the words. However, although an even number of odd items would be preferable, this is not considered to have affected the learners' responses, as this only happened with ten out of 80 triads. In addition, an effort was made so that every triad contained male and female voices, also randomising the frequency with which the male or female was the odd item out.

The discrimination test was created in TP as five different activities. The first was a training session like the one in study 1. The aim was to familiarise participants with the aural oddity task. It included 10 triads with minimally paired words featuring non-target sound contrasts that were clearly phonetically distant (e.g. *pat-pat-pit; ping-pong-ping; put-pit-put*), also including catch triads (e.g. *ping-ping-ping*). The second addressed the four vowel contrasts explained above, and the third, fourth and fifth addressed the /s – z/ contrast in initial, medial and final position respectively (Figure 60).

(1) Discrimination task

a. Activity 1: training session with non-target stimuli
b. Activity 2: /æ-ʌ/, /æ-ɑː/, /ʌ-ɑː/, /ə vs strong version/
c. Activity 3: /s – z/ in initial position
d. Activity 4: /s – z/ in medial position
e. Activity 5: /s – z/ in final position

(2) Identification task

a. Activity 1: vowel sounds with only one possible target sound
b. Activity 2: vowel sounds words with various possible target sounds
c. Activity 3: /s – z/

Figure 60. Structure of the perception pre- and post-tests for study 2

Stimuli for the identification test consisted of 120 tokens (excluding individual sounds in the first activity). Each target sound was featured in 20 tokens, 10 familiar and 10 novel. As in study 1, familiar items were selected after an inspection of the number of words featuring each sound in the training stimuli as well as their frequency of occurrence.

The main criteria for stimuli selection were the frequency of appearance of that word during the trial run mentioned above and the orthographic representations featuring the target sounds in those words.[97] Novel stimuli presented the target sounds in different positions and spelt in different ways. Some of the novel stimuli for schwa were purposefully tested in part 3 of the test so that students saw the words in writing. This was done in order to test the learners' identification of the sound despite the influence of spelling, given the learners' tendency to substitute schwa for the letter that coincides with the phonemic symbol for Spanish (e.g. in *December* for /e/, in *hospital* for /a/, in *family* for /i/ – see Monroy, 2001 for examples). For the novel words featuring the /s – z/ contrast, two had the contrast in initial, four in medial and four in final position. A list with the stimuli for the identification task for every sound can be found in Appendix 19.

Production

The production test for this study shared the same format and procedure as that in study 1, with the exception of the accent-mimicry task used in that study used to test perception of /b d g/. Production pre- and post-tests consisted of the same three parts aimed at assessing the participants' production of the target sounds in three different modes: imitative, controlled, and spontaneous production (see section 3.1.3). Stimuli for these three tasks are explained in the same order in which the tasks were conducted, starting with the sentence-reading task, followed by the timed picture-description, and finally, the imitation task. The imitation task was conducted last in order to avoid possible training effects.

For the sentence-reading task, a list of the most commonly occurring words in the app for each target sound was compiled in order to test the learners' production of 'familiar' words (i.e. those that appeared during training). Each sound was featured in 10 familiar words from the ones that occurred most often in the app. As in the stimuli for the identification task, the word *glasses* only appeared in activity 2 (i.e. no audio input was offered for this word – at least in the first 1000 stimuli of the trial run explained above),

---

[97] In those cases where several words had the same number of occurrence (NO) during training. The selection was made based on their frequency of occurrence (FOC) in the British National Corpus (http://www.natcorp.ox.ac.uk/). For example, for stimuli featuring /æ/, there six words that appeared four times. Words like *chat* (FOC = 1266), or *salad* (FOC = 1103) were excluded in favour of words like *have* (FOC = 460626), *back* (FOC = 96939), *black* (FOC = 23830, or *bad* (FOC = 14903). For stimuli featuring /ə/, there were also six words that appeared four times. In this case, *hairdresser* (FOC = 395) and *secretary* (FOC = 15359), were excluded in favour of words like *second* (FOC = 40664), *October* (FOC =10529), *sugar* (FOC =3691), or *pilot* (FOC =3113). Even though an attempt was made to include high-frequency words such as *second*, or *October*, less frequent words such as *pilot* or *sugar* were included instead of *secretary* in order to offer a wider variety of orthographic representations (/ə/ represented by the spelling <o>).

but it was included anyway because there were no more words featuring /ɑː/ in the app. Even though it only appeared in the testing phase (i.e. the activity without audio where users have to recall what they have learnt from activity 1), this should make learners aware that the word *glasses* is pronounced with /ɑː/ in SSBE.

Additionally, five novel words with the sound in different contexts and spelt in different ways were included in order to test possible generalisation gains. Given the high number of target sounds in this study, novel words were limited to five tokens per sound. As in study 1, generalisation in these studies focuses on transfer of potential learning gains to new orthographic contexts in lexical items that do not appear in training. As pointed out above, English vowels can be represented by a range of spellings, which may be a further complication for students, as they have to learn to perceive and produce the sound, but also learn when the sound is required.

A list with all the stimuli used for the sentence-reading task and their carrier sentences can be found in Appendix 20. As in study 1, items featuring the target sounds were included in the carrier sentence '*The ____ is in/on the ____*' after Flege and Hammond (1982). This guaranteed that every speaker pronounced words in the same (controlled) phonetic context, with the target sounds in accented position – except for schwa, which appeared in unaccented contexts. This was done with the intention of eliciting vowels in contexts where they should appear in their strong form.[98] Nonetheless, the carrier sentence was slightly modified in some cases in order to keep a reasonable number of examples for each target sound, yet limiting test duration to no more than 10-12 minutes.

As regards the timed picture-description task, given the demanding nature of FL vowel sounds explained above (in terms of articulation and different orthographic representations), words were limited to familiar tokens (Table 25). Each sound was featured in four words that appeared in training. Learners were presented a total of 15 pictures that contained three word prompts including target items from different sounds as well as distractors.

[98] Sounds in stressed syllables are produced with greater muscular effort and vowels have clear quality and full length, as opposed to unstressed syllables where vowels may have a centralised quality and are comparatively shorter (Collins & Mees, 2009)

Table 25. Stimuli for the timed picture-description task (study 2)

	/æ/		/ʌ/		/ɑ:/		/ə/		/z/
1	stamp	5	cousin	9	aunt	13	famous	17	music
2	capital	6	uncle	10	car	14	Africa	18	buys
3	actor	7	young	11	dark	15	picture	19	flies
4	black	8	son	12	dance	16	dangerous	20	museum

Finally, the imitation task consisted of a total of 20 stimuli obtained from the same compilation used in the discrimination task, featuring vowel sounds in different phonetic contexts. Stimuli for schwa included tokens that often present fossilised pronunciations as a result of their spelling (e.g. <er> or <e> in *manner* or *problem*, commonly mispronounced with Spanish /e/, <a> in *hospital* often mispronounced with /a/, <o> in *prison*, often mispronounced with /o/ – see Monroy, 2001 for examples). A list with all the words used in the imitation task can be found in Table 26.

Table 26. Stimuli for the imitation task in study 2.

	/æ/		/ɑ:/		/ʌ/		/ə/		/z/
1	hat	5	heart	9	cup	13	manner	17	zeal
2	bag	6	barn	10	buck	14	hospital	18	z
3	cap	7	park	11	huff	15	problem	19	fuzzy
4	track	8	chart	12	cluck	16	prison	20	prize

f) Procedure for the evaluation of production stimuli

The procedure adopted for the evaluation of production data was the same as that in study 1. Audios from the production pre- and post-tests were edited individually with Audacity and saved as separate files for their subsequent evaluation by the expert judges introduced above (10 m. approx. x 62 participants[99] x 2 testing times = ca. 20.6 hours of audio). As in the podcasts study, despite the fact the total number of items should yield 12,650 tokens (Table 27), learners occasionally skipped words in either the pre- or post-tests. The final number of tokens was 11,160.

[99] Even though the total number of participants in this study was 63, there was one participant who did not complete the final production test. Thus, the data cannot be taken into consideration for the analysis, nor are they considered here.

Table 27. Number of tokens obtained per target item, task, in total, and for all participants (study 2)

Sentence-reading task							
/æ/	**/ʌ/**	**/ɑː/**	**/ə/**	**/z/**	**Total**	**(pre- and post-tests)**	**x 62 participants**
15	15	15	15	15	75	150	9300
Picture-description task							
/æ/	**/ʌ/**	**/ɑː/**	**/ə/**	**/z/**	**Total**	**(pre- and post-tests)**	**x 62 participants**
4	4	4	4	4	20	40	2480
Imitation task							
/æ/	**/ʌ/**	**/ɑː/**	**/ə/**	**/z/**	**Total**	**(pre- and post-tests)**	**x 62 participants**
4	4	4	4	4	20	40	2480
				Total	115	230	14260
Total after missing audios							11160

As in study 1, a total of 20 tokens were evaluated twice in order to calculate intra-rater reliability. This time, since there were five target sounds for production, tokens included four tokens from each sound, with the same number of pre- and post-test productions, always from different speakers (see Table 28).

Table 28. Stimuli selection for intra-rater reliability for study 2

Target sound	Tokens	Participant	Pre- or post-test
/æ/	stamp	1	pre
	capital	10	post
	happen	7	pre
	actor	14	post
/ʌ/	hundred	24	pre
	son	18	post
	stomach	30	pre
	under	26	post
/ɑː/	argue	62	pre
	aunt	63	post
	car	22	pre
	dark	38	post
/ə/	dangerous	50	pre
	dinner	45	post
	October	100	pre
	pilot	101	post
/z/	flies	84	pre
	buys	90	post
	reads	96	pre
	museum	78	post
Summary	20 tokens	20 participants	10 pre/10 post

5.3.3 Results

Following the same procedure adopted in section 4.2.3, the results obtained in the five tasks explained above will be presented in the following order: two perception (discrimination and identification) and three production (imitation, sentence-reading and timed picture-description) tasks.

As in study 1, the total scores for each target aspect will be presented first, dealing initially with the differences between familiar and novel stimuli and then among tasks. As explained in section 5.2.2a, there were three groups in this study. Two (G1 and G3) acted as experimental groups and G2 acted as control during training. However, after G1 and G3 finished training, G2 started to receive instruction, therefore acting as experimental group too. Hence, it was expected that G1 and G3 improved their perception and production of the target sounds from pre- to post-test as compared to G2, and G2 was expected to improve their perception performance from the post-test to the delayed post-test.

The differences between pre- and post-tests scores by the experimental (G1 and G3) and control groups (G2) were analysed with two-way mixed measures ANOVAs. When a significant effect was found in the interaction between the time and group variables, post-hoc Holm tests were conducted in order to find out for which of the groups interaction reached significance. In order to measure improvements in G2's performance once they received instruction, paired T-tests were used in order to compare the improvement they made when acting as control group (from pre- to post-test) to the one they made when acting as experimental (from post-test to delayed post-test). Thus, this section follows the same pattern, addressing differences between experimental (G1 and G3) and control group (G2) first, and then, when applicable, exploring possible differences between experimental groups (G1 and G3); finally, the section presents the degree of improvement obtained by G2 from the post-test to the delayed post-test.[100]

Perception

The discrimination task consisted of 95 triads, including training 10 triads aimed at familiarising learners with the procedure and 5 distractors. Therefore, scores for the discrimination task were obtained by counting the number of correct responses out of 75

[100] As with the tests carried out in study 1 (section 4.2.3), Shapiro-Wilk tests were conducted in order to test the normality of data and Barlett or Levene tests to test whether samples have equal variances. For the T-tests, when Shapiro-Wilk tests indicated that the data were not normally distributed, Wilcoxon-Signed-Rank Tests were used (a non-parametric test similar to the T-test).

triads (see section 5.2.2d). The native speakers' scores on this task were 67 (89.3%), 62 (82.7%) and 69 (92%). The expert judges' scores in this task were 70 (93.3), 64 (85.3%) and 66 (88%).

The ANOVA performed with the total mean scores from pre- and post-tests for the discrimination task revealed a significant effect of the time variable ($F(1,56) = 14.9$, $p = 0.000293 < 0.05$), but no interaction effects between group and time ($F(1,56) = 0.123$, $p = 0.883 > 0.05$). That is, there were significant increases in the learners' scores from pre- to post-test, but all groups improved similarly (Figure 61). Similarly, the paired T-test comparing the post-test and delayed post-test scores for G2 showed that improvements did not reach significance for G2 either ($t(22) = -0.83$, $p = 0.415 > 0.05$).

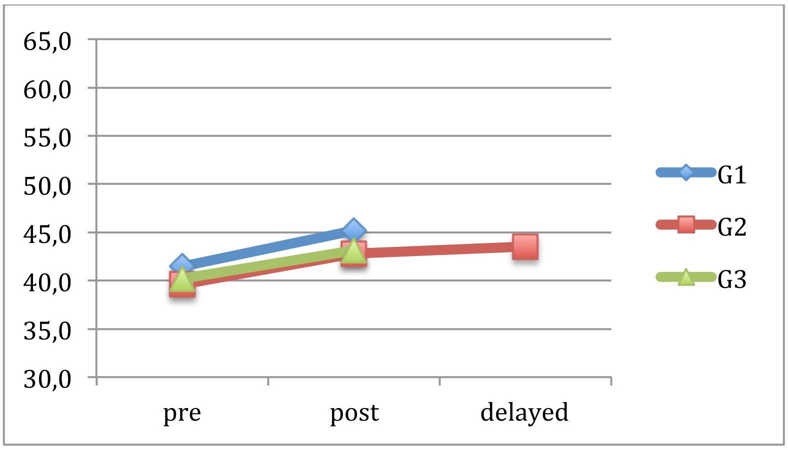


Figure 61. Degree of improvement made in the discrimination task

Additionally, the results obtained for each pair of sounds tested in the triads were also analysed. As explained in section 5.2.2d, there were 10 triads for each pair of target sounds (/æ – ʌ/, /æ – ɑː/, /ʌ – ɑː/, /s – z/ and the weak and strong versions of words containing schwa in their reduced form) and 25 catch triads with 5 for each target sound except for schwa. In line with the results obtained for the total scores for the discrimination task, no significant differences were found among groups. There was a significant effect of the time variable ($F(1,56) = 15.5$, $p = 0.000235 < 0.05$) for the results obtained for the /ʌ – ɑː/ contrast and in the catch triads for /ʌ/ ($F(1,56) = 6.29$, $p = 0.015 < 0.05$) and /z/ ($F(1,56) = 4.73$, $p = 0.0338 < 0.05$), but no interaction effects were found between group and time. In other words, the improvement made by the three groups was similar.

The identification task consisted of 120 tokens aimed at measuring the learners' identification of the target sounds. The expert judges' scores in this task were 120 (100%), 114 (95%) and 105 (87%).

The results for the total scores in this task show that there was a significant effect of time ($F(1,52) = 139$, $p = 2.49e\text{-}16 < 0.05$), and a significant interaction effect between time and group ($F(2,52) = 16.7$, $p = 2.5e\text{-}06 < 0.05$). The results from the Holm post-hoc test indicate that the interaction effect was significant both for G1 ($p = 0.000122$) and for G3 ($p = 0.019$). Furthermore, an ANOVA was performed in order to compare the results by G1 and G3 (the first two groups acting as experimental), but no significant differences were found between groups ($F(1,31) = 0.927$, $p = 0.342 > 0.05$). In other words, G1 and G3 improved significantly from pre- to post-test, while G2's scores remained stable. The average scores for G1 increased from 65.1 (16.4) points (54.3%) to 84.8 (13.5) points (70.7%), with an improvement of 19.7 points (16.4%). The scores for G3 increased from 53.7 (14.1) points (44.8%) to 77.4 (19) (64.5%), showing an improvement of 23.7 points (19.8%). Nonetheless, the scores obtained by G2 increased by 6.3 points (5.3%), with a mean score of 68.7 (12.5) points (57.3%) in the pre-test and 75 (15.1) in the post-test (62.5%). The data reveal a significant effect of treatment, which finds further support in the significant improvement made by G2 once they received instruction ($t(17) = -2.357$, $p = 0.03070282$). G2's mean scores increased by 16.2 points (13.5%), with 75 points (15.1) in the post-test (62.5%) and 91.2 (11.6) in the delayed post-test (76%) – see Figure 62.

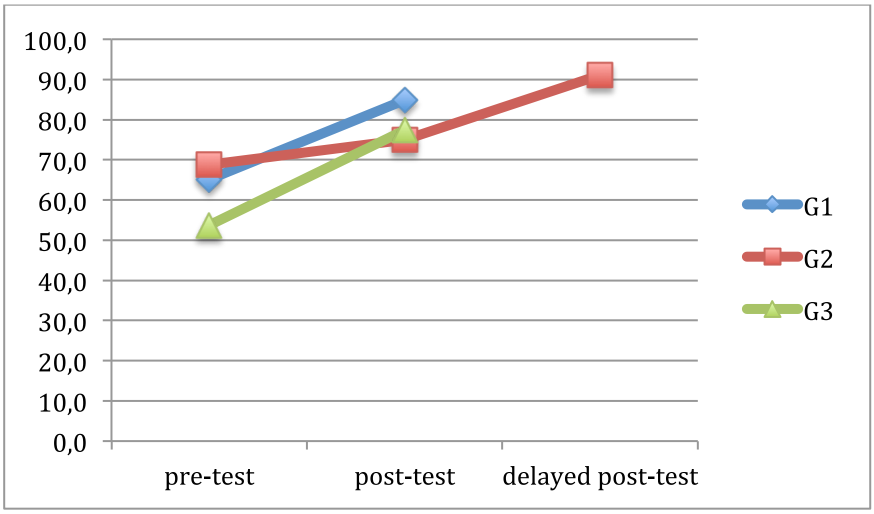


Figure 62. Degree of improvement made in the identification task showing average scores on the vertical axis and testing time on the horizontal axis

In order to investigate possible differences of the impact of instruction on each target sound, the scores obtained in stimuli featuring each sound were calculated. As explained above, each sound was featured in 20 items (10 familiar and 10 novel). Hence, the maximum score for each sound was 20. The results of the three groups for each target sound are presented in Table 29.

Table 29. Mean scores and degree of improvement (imp) per group for each sound in the identification task

					/æ/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	9.0 (4)	44.8	12.3 (4.3)	61.7	3.4	16.9				
G2	9.4 (3)	46.8	11.3 (3.7)	56.7	2.0	9.9	14.0 (3.4)	70.0	2.7	13.3
G3	8.1 (3.2)	40.5	12.2 (4)	61.1	4.1	20.6				
					/ʌ/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	9.6 (3.8)	48.1	14.5 (3.9)	72.5	4.9	24.4				
G2	11.8 (3)	59	12.9 (4)	64.6	1.1	5.6	16.3 (2.5)	81.7	3.4	17.2
G3	9.5 (3.4)	47.5	12.9 (3.6)	64.4	3.4	16.9				
					/ɑː/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	8.3 (3.1)	41.3	11.2 (2.2)	55.8	2.9	14.6				
G2	8.9 (3.1)	44.5	9.1 (3.1)	45.7	0.2	1.2	12.1 (3.1)	60.4	3.0	14.8
G3	7.6 (3.7)	38	10.3 (4.4)	51.7	2.7	13.7				
					/ə/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	11.0 (4.7)	54.8	15.0 (3.1)	74.8	4.0	20.0				
G2	12.1 (4.4)	60.4	13.3 (4.8)	66.5	1.2	6.1	16.7 (2.1)	83.3	3.3	16.7
G3	4.9 (2.9)	24.5	11.7 (4.4)	58.3	6.8	33.8				
					/s/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	15.4 (3.9)	76.9	17.0 (2.5)	85.2	1.7	8.3				
G2	14.8 (2.9)	74.2	16.0 (3.2)	80.2	1.2	6.0	16.8 (2.9)	83.9	0.7	3.7
G3	14.1 (4)	70.5	16.9 (3.4)	84.4	2.8	13.9				
					/z/					
	pre-test	**%**	**post-test**	**%**	**imp**	**%**	**delayed**	**%**	**imp**	**%**
G1	12.0 (3.4)	59.8	14.8 (3.7)	74.2	2.9	14.4				
G2	11.9 (3.5)	59.6	12.3 (3.9)	61.3	0.3	1.7	15.4 (2.6)	77.2	3.2	15.9
G3	9.5 (5)	47.5	13.4 (5.6)	67.2	3.9	19.7				

The results obtained from the ANOVAs comparing pre- and post-test scores for all groups revealed that there were significant differences in the degree of improvement made between experimental and control groups for /ʌ/, /ɑː/, /ə/, and /z/, as indicated by an interaction effect between time and group. However, Holm post-hoc tests showed that these interaction effects did not reach significance for the two experimental groups for every sound. Instruction was significant for both experimental groups (G1 and G3) for

/ə/. Nonetheless, the improvements in /ʌ/, /ɑː/ and /z/ only reached significance for G1. The p-valuess from Holm post-hoc tests can be found in Table 30 below. These are only included when the ANOVA revealed a significant interaction between group and time. As regards the scores for G2 once they received instruction, the improvements made when they acted as experimental group did not reach significance as compared to their performance as control group.

In addition, a significant effect of the group variable was found when comparing all groups' performance with /ə/ ($F(1,31) = 14.3$, $p = 0.000676 < 0.05$), which indicates that G3's mean scores for /ə/ were considerably lower than those by G1 and G2 at the beginning of the study.

Table 30. *p-values*s obtained from the ANOVAs and Holm post-hoc tests comparing pre- and post-test scores for experimental (G1 and G3) and control groups (G2) for each sound in the identification task

Sound	Interaction group:time (pre-post)	Holm post-hoc tests
/æ/	$F(2,52) = 1.88$, $p = 0.1629 > 0.05$	
/ʌ/	$F(2,52) = 6.28$, $p = 0.0036 < 0.05$ *	G1 ($p = 0.000208$)*; G2 ($p = 0.233$); G3 ($p = 0.106$)
/ɑː/	$F(2,52) = 5.24$, $p = 0.0084 < 0.05$ *	G1 ($p = 0.00148$)*; G2 ($p = 0.853$); G3 ($p = 0.429$)
/ə/	$F(2,52) = 6.35$, $p = 0.0034 < 0.05$ *	G1 ($p = 0.00329$)*; G2 ($p = 0.337$); G3 ($p = 0.00329$)*
/s/	$F(2,52) = 1.23$, $p = 0.2996 > 0.05$	
/z/	$F(2,52) = 5.03$, $p = 0.0101 < 0.05$ *	G1 ($p = 0.0221$)*; G2 ($p = 0.694$); G3 ($p = 0.277$)

The improvements made by participants when acting as experimental groups indicate that the most substantial gains occurred for /ə/, followed by /ʌ/, /æ/, /z/, /ɑː/ and finally, /s/. The learners' mean scores in the pre-test were mostly below 50 per cent of the total score, with the exception of G2's scores for /ʌ/ (59%), G1 and G2's scores for /z/ (around 59%) and /ə/ (between 54 and 60%), or all groups' scores for /s/ (which were all above 70% of the maximum score in the pre-test). Overall, the lowest pre-test mean scores were found fo /æ/ (40-47% range), /ʌ/ (47-59% range) and /ɑː/ (38-44.5% range), although G3 obtained an even lower score for /ə/ (24.5%). However, the participants' post-test scores were much higher. G1's mean score was above 70 per cent in several target sounds and G2's mean scores after receiving instruction were higher than 70 per cent for most sounds. G3's post-test scores were comparatively lower, only surpassing the 70 per cent performance for /s/.

In order to find out whether learners could generalise improvements to novel words, the mean scores for each group in familiar and novel words were explored separately (Table 31). The scores obtained for novel stimuli comparing all groups' performance in pre- and post-tests were analysed with an ANOVA. The results revealed a significant effect of time ($F(1,52) = 47.1$, $p = 8.21e\text{-}09 < 0.05$) and a significant interaction effect between time and group ($F(2,52) = 5.17$, $p = 0.00896 < 0.05$). The results from the Holm post-hoc test revealed that interaction reach significance only for G1 ($p = 0.0291$*). The *p-values*s for G2 and G3 were higher than 0.05 ($p = 0.345$ for G2 and $p = 0.157$ for G3). Likewise, no significant differences were found in the results from the paired T-test analysing G2's performance when acting as experimental.

Table 31. Mean scores for the identification task: familiar and novel stimuli

Familiar stimuli					
Max (60)	**pre-test**	**post-test**	**imp**	**delayed**	**imp**
G1	32.4 (8.4)	46.0 (8.4)	13.6	-	-
G2	33.8 (6.1)	37.6 (8.8)	3.8	49.6 (5.2)	12
G3	27.7 (6.8)	43.1 (8.8)	15.4	-	-
Novel stimuli					
Max (60)	**pre-test**	**post-test**	**imp**	**delayed**	**imp**
G1	32.8 (8.7)	38.8 (6.8)	6.1	-	-
G2	34.9 (7.4)	37.4 (7.3)	2.5	41.7 (6.9)	4.3
G3	26.0 (7.7)	34.3 (10.7)	8.3	-	-

Note: The table includes the average scores for the identification task at pre-test, post-test and delayed post-test (delayed) with standard deviations in parentheses, the degree of improvement from one testing time to the next (imp), as well as the maximum scores for each block of stimuli (max)

The improvement achieved in familiar items show a similar trend than the data for the whole set of stimuli. G3 made the most substantial improvement, with 15.4 points (25.7%) from pre- to post-test, followed by G1 with 13.6 points (22.7%). The control group improved by only 3.8 points (6.3) from pre- to post-test, revealing a comparatively smaller improvement. However, their average scores after training increased by 12 points (20%), therefore providing further support for the positive effect of training despite not reaching significance (Table 31).

Table 32. Mean scores and degree of improvement made for each sound on familiar and novel stimuli

	Familiar					Novel				
/æ/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	4.8	6.9	2.1			4.2	5.5	1.3		
%	47.5	68.8	21.3			42.1	54.6	12.5		
G2	5	6	1	8	1.9	4.3	5.3	1	6	0.7
%	50.4	60.4	10.1	79.6	19.1	43.2	53	9.8	60.4	7.4
G3	4.4	7.3	2.9			3.7	4.9	1.2		
%	44	73.3	29.3			37	48.9	11.9		
/ʌ/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	4.5	7.5	3			5.1	7	1.8		
%	45	75.4	30.4			51.3	69.6	18.3		
G2	5.7	6.3	0.7	8.6	2.3	6.2	6.6	0.4	7.7	1.2
%	56.9	63.5	6.6	86.1	22.6	62	65.7	3.7	77.4	11.7
G3	4.5	6.7	2.2			5	6.2	1.2		
%	45	66.7	21.7			50	62.2	12.2		
/ɑː/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	3.6	6.3	2.8			4.7	4.8	0.2		
%	35.8	63.3	27.5			46.7	48.3	1.7		
G2	3.9	4.3	0.4	6.7	2.4	5.1	4.9	-0.3	5.4	0.5
%	38.8	42.6	3.8	67	24.3	51.2	48.7	-2.5	53.9	5.2
G3	4.1	5.7	1.6			3.5	4.7	1.2		
%	41	56.7	15.7			35	46.7	11.7		
/ə/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	5.7	8.5	2.8			5.3	6.5	1.2		
%	57.1	85	27.9			52.5	64.6	12.1		
G2	6.5	7.1	0.7	9.1	2	5.7	6.2	0.5	7.6	1.4
%	64.6	71.3	6.7	90.9	19.6	56.8	61.7	4.9	75.7	13.9
G3	2.7	7.1	4.4			2.2	4.6	2.4		
%	27	71.1	44.1			22	45.6	23.6		
/s/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	8.2	9.2	1			7.2	7.9	0.7		
%	82.1	91.7	9.6			71.7	78.8	7.1		
G2	7.5	8.1	0.7	9.1	1	7.5	7.9	0.4	7.7	-0.3
%	74.6	81.3	6.7	91.3	10	74.8	79.1	4.3	76.5	-2.6
G3	7.9	9	1.1			6.2	7.9	1.7		
%	79	90	11			62	78.9	16.9		
/z/	**pre**	**post**	**imp**	**delayed**	**imp**	**pre**	**post**	**imp**	**delayed**	**imp**
G1	5.6	7.6	2			6.3	7.3	0.9		
%	56.3	75.8	19.6			63.3	72.5	9.2		
G2	5.4	5.7	0.3	8.1	2.4	6.5	6.6	0	7.3	0.8
%	54.2	57	2.7	80.9	23.9	65.2	65.7	0.5	73.5	7.8
G3	4.1	7.3	3.2			5.4	6.1	0.7		
%	41.0	73.3	32.3			54.0	61.1	7.1		

A closer inspection of the data for familiar and novel stimuli reveals that despite the fact that a significant effect was found for G1 when considering the total performance for novel stimuli across sounds, no significant interactions between time and group were found in novel stimuli for individual sounds for G1. The only significant difference in improvements among groups for individual sounds in novel stimuli was the one by G2 for /ɑː/ when acting as experimental ($t(19) = -2.336$, $p=0.03059915$). They improved by 0.5 points (5.2%) from the post-test to the delayed post-test, as compared to the decrease (0.3 points) in their average performance when acting as control.

The above indicates that, overall, the differences in the improvement made among groups in novel stimuli only reached significance when the former were considered collectively. As can be seen in Table 32, the degree of improvement made in novel stimuli was in most cases considerably lower than the one in familiar stimuli. There are, however, exceptions in which the difference is not so large, as in G2's improvements for /æ/or /ə/ in novel stimuli when acting as control, or the three groups' improvement for /s/.

Production

With regard to the production tasks, some preliminary considerations regarding the stimuli analised are offered here before the analysis itself. Four items were eliminated from the production stimuli and were not considered in the analyses. These were *have*, from the familiar stimuli for /æ/, and *run* from the novel stimuli for /ʌ/ in the sentence-reading task; and *son* from the stimuli for /ʌ/ and *buys* from the stimuli for /z/ in the timed picture-description task. The maximum score for each sound per task can be found in Table 33.

In order to make testing stimuli fit in the smallest number of sentences possible, the researcher tried numerous possible combinations for carrier sentences. However, in the last arrangement, he mistakenly placed *have* and *run* in positions that did not yield the desired pronunciations for measuring those sounds. *Have* was embedded in the sentence *I have a flat*. The sentence was intended to elicit the strong version of *have* (i.e. with /æ/ rather than schwa), but the vast majority pronounced the reduced version. Hence, the stimulus was not considered for the analysis. As for the word *run*, in one of the trials for the possible carrier sentences, the researcher mistakenly placed *run* (past participle) in a position that required the past simple form of the verb (i.e. *ran*). Even though learners were reading the sentence and could have interpreted it as *run* (i.e. /ʌ/), the stimulus was also omitted from the analysis.

Additionally, two stimuli from the timed-picture description task were also omitted from the analysis (i.e. *countries* and *buys*). The researcher had the initial intention of including both novel and familiar words in this task and had compiled a larger list of stimuli for this part of the test. However, because adding novel words to this task would increase the duration of the tests considerably, novel words were finally omitted. Since pictures and prompts for the task were already created, when reducing the number of novel stimuli, the researcher mistakenly deleted the slide featuring the word *son* and left one featuring the novel word *countries*. Nevertheless, in order to keep the task as homogeneous as possible, the word *countries* was not considered for the analysis, as it would be the only novel word. The second was the word *buys*. It was included in the task because it is one of the most common words among the training stimuli. The picture on the slide showed a man at a music shop and the prompts were *buy, music, download*. The sentence the researcher had in mind when writing the prompts was some kind of general statement defining the man's attitude to music: *He buys music, he doesn't download music.* Nevertheless, most learners described the picture by defining what was happening at the moment, making sentences such as *He is buying music, he doesn't download it*. Thus, the word *buys* was not considered in the analysis either.

Finally, it should be noted that given the demanding nature of stimuli rating, it was impossible to complete the evaluation of the learners' production of /ɑː/ in time for the analysis. The ratings took place between April and July but it was impossible to meet with the other two judges after that date. Thus, the production data for /ɑː/ are not considered in the analyses either.

The interrater reliability as measured by the Fleiss' Kappa test was found to be 0.979, which can be interpreted as 'almost perfect agreement' (0.81-1.00 range). Furthermore, as explained above, intra-rater reliability, was measured by playing a selection of 20 extra items that had already been assessed. The intra-rater reliability was perfect in this study, as judges always rated the extra items with the same score they had given them before. Thus, no extra tests were conducted. It is important to point out that, even though the total number of tokens should be 14,260 (see section 5.2.2f above), the final number of tokens for evaluation was 11,160. This was due to the fact that learners sometimes skipped words in either the pre- or post-tests, but also because the stimuli for /ɑː/ were not evaluated.

Table 33. Maximum scores for each sound the production tasks: Imitation task (IT), sentence-reading task (SRT) and timed picture-description task (TPDT)

Stimuli	IT	SRT	TPDT	Total
/æ/	4	14 (9 familiar/5 novel)	4	22
/ʌ/	4	14 (10 familiar/4 novel)	3	21
/ə/	4	15 (10 familiar/5 novel)	4	23
/z/	4	15 (10 familiar/5 novel)	3	22

The ANOVA performed with the total scores from the production tests across sounds revealed a significant effect of time ($F(1,59) = 60.7$, $p = 1.22\text{e-}10 < 0.05$) and a significant interaction effect between group and time ($F(2, 59) = 5.1$, $p = 0.00902 < 0.05$). That is, there were significant differences in the improvement made among groups for production. G1's mean scores increased from 15.9 (9.2) points in the pre-test to 24.9 (13) points in the post-test, showing an improvement of 9 points (10.2%). The mean scores by participants in G3 increased by 8.7 points (9.8%), with a mean score of 19.1 (15.8) points (21.7%) in the pre-test and 27.8 (23.1) in the post-test (31.6%). Finally, G2's scores increased by 3.7 points (4.2%) from pre- to post-test, with a mean score of 14 (9.7) points in the pre-test (16%) and 17.7 (12.3) in the post-test (20.2%). The data show that the overall production improvement made by the two experimental groups was superior to that by G2. Nevertheless, the results from the Holm post-hoc tests revealed that G1 was the only group for which differences reached significance (G1 $p = 0.0182$*; G2 $p = 0.447$; G3 $p = 0.447$). Moreover, it should be noted that the mean scores by participants in the three groups were far from the maximum production score (88 points) – see Figure 63.

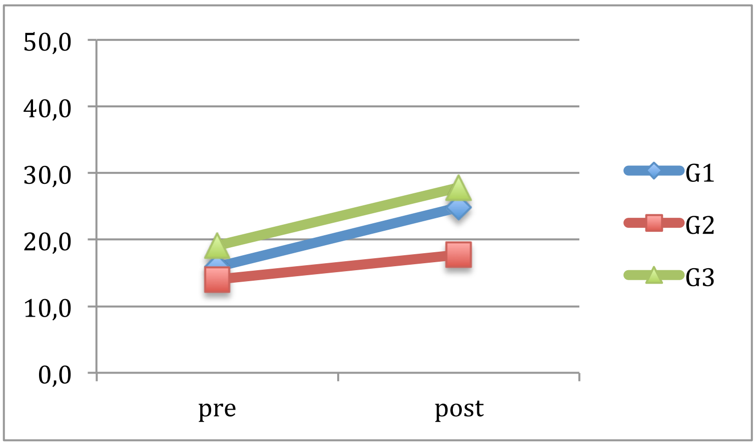


Figure 63. Average scores and degree of improvement achieved in production across groups

The analyses carried out with the scores for individual sounds revealed significant effects of the time variable for all the target sounds. However, significant interaction effects were only found for /æ/, /ʌ/ and /ə/. That is, the three groups made similar improvements for /z/, but there were significant differences among groups in the improvement for /æ/, /ʌ/ and /ə/ (see Table 34). The Holm post-hoc tests showed that the significant interaction effects were only found for G1. Despite the fact that the two experimental groups made similar improvements between testing times (see Figure 64), the improvement made by participants in G3 did not reach significance.

Table 34. Results from the ANOVAS comparing all groups' performance for the total production scores for each sound from pre- to post-test

Sound	Time variable	Interaction between time and group
/æ/	F (1,59) = 35.3, p = 1.64e-07 < 0.05*	F (2,59) = 3.88, p = 0.0262 < 0.05*
/ʌ/	F (1,59) = 38.9, p = 5.17e-08 < 0.05*	F (2,59) = 3.76, p = 0.0291 < 0.05*
/ə/	F (1,59) = 23.5, p = 9.29e-06 < 0.05*	F (2,59) = 3.19, p = 0.0483 < 0.05*
/z/	F (1,59) = 10.1, p = 0.00239 < 0.05*	F (2,59) = 0.43, p = 0.6516 > 0.05

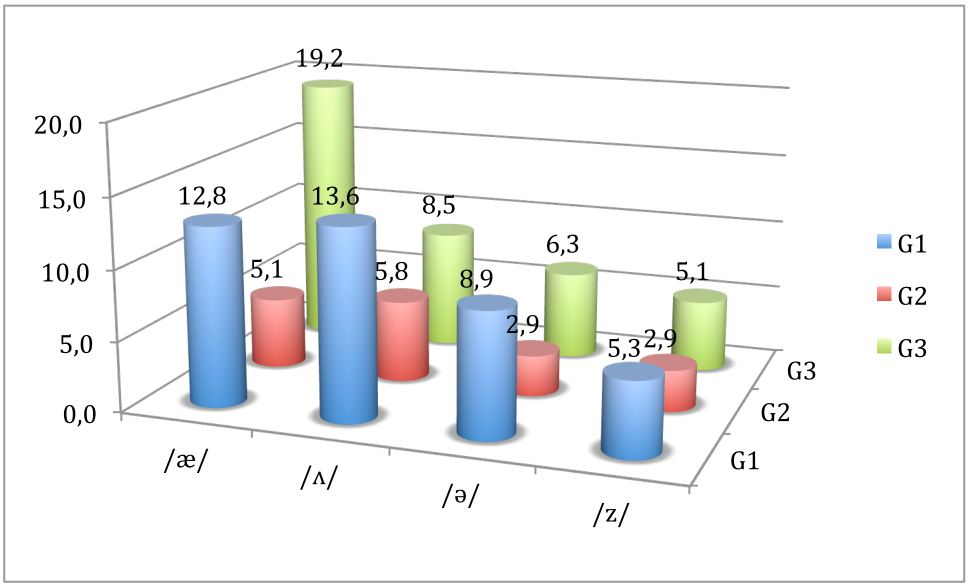


Figure 64. Percentage of improvement made each sound for G1, G2 and G3 considering the maximum score possible for each sound

Table 35 presents the data for individual sounds, including the mean scores in pre- and post-tests for all groups, standard deviations (SD) and the degree of improvement (imp) made. Some of the average scores present sizeable standard deviations, sometimes even higher than the mean. This implies that the learners' individual scores vary greatly, with some learners obtaining considerable improvements from instruction and others remaining relatively stable. The individual scores obtained by participants for each sound can be found in Appendix 21.

Table 35. Mean scores, standard deviations and degree of improvement made for each sound in the production tasks (across tasks)

	/æ/			/ʌ/			/ə/			/z/		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	2.7 (2.6)	5.5 (4.9)	2.8	6 (2.4)	8.8 (3.2)	2.8	4.9 (4.4)	7 (4.9)	2	2.4 (4.7)	3.7 (5.5)	1.3
%	12.1	24.8	12.8	28.4	41.9	13.6	21.4	30.3	8.9	10.8	16.6	5.8
G2	2.6 (3.9)	3.7 (4.7)	1.1	6.3 (2.7)	7.5 (3.2)	1.2	4.0 (4.3)	4.7 (4.5)	0.7	1.2 (1.5)	1.9 (2.6)	0.7
%	11.6	16.7	5.1	29.8	35.6	5.8	17.4	20.3	2.9	5.6	8.8	3.2
G3	2.2 (3.2)	6.4 (6.1)	4.2	7.3 (4.5)	9.1 (5.4)	1.8	6.2 (6.9)	7.7 (7.3)	1.4	3.3 (5)	4.6 (7.8)	1.2
%	10.1	29.3	19.2	34.9	43.4	8.5	27.1	33.3	6.3	15.2	20.7	5.6

Focusing on G1, their scores for /æ/ show that there were 10 participants who made moderate or no improvements from pre- to post-test or even obtained lower scores in the post-test. Nine participants improved in one (4.5% of improvement considering the total score for this sound) to four items (18.2% of improvement), and eight participants improved five or more items (from 22.7% of improvement to 54.5%). However, the most noticeable improvement in the control group (G2) was the one by participants 22 and 53. They improved six items (27.3%) from pre- to post-test, which is half of the improvement made by the best participant in G1. The mean rate of improvement for this group is affected by these two participants, as the vast majority improved in two (9.1%) or fewer items. As for G3, one participant showed substantial gains from instruction, with an improvement of 13 points (59.1%), two participants improved 7 points (31.8%) and the rest improved five (22.7%) or fewer points.

With regard to the imitation task, the developments in the learners' production was generally very similar. The analysis of the total scores for this task revealed a significant effect of the time variable ($F(1,59) = 27.8$, $p = 2.02e\text{-}06 < 0.05$), but no significant interaction effects were found between time and group ($F(1,59) = 1.387$, $p = 0.257 > 0.05$). That is, even though there were significant improvements from pre- to post-test, all groups progressed similarly.

The ANOVAs analysing the participants' performance for each sound revealed significant effects of the time variable for /æ/ ($F(1,59) = 12.5$, $p = 0.000815 < 0.05$), /ʌ/ ($F(1,59) = 17.6$, $p = 0.0000926 < 0.05$) and /z/ ($F(1,59) = 4.81$, $p = 0.0322 < 0.05$).

Nevertheless, /æ/ was the only sound that showed a significant interaction effect between time and group ($F(2,59) = 3.37$, $p = 0.0412 < 0.05$). In line with the scores from perception tests, Holm post-hoc tests revealed that G1 was the only group for which interaction reached significance.

Overall, the improvements made by the three groups in the imitation task were very similar across sounds. Table 36 shows the mean scores for each group at pre- and post-tests including the degree of improvement. As can be seen, the improvement made by the experimental groups was greater than the one by G2 in /æ/ and /ʌ/. However, the differences between G2 and G3 were very subtle for the latter sound. For /ə/ and /z/, the improvements made by G2 were very similar or even greater than the ones by the experimental groups.

The pre-test scores show that /æ/ was the most problematic sound for participants in G1 and G3, with mean performances of 15.4 per cent (G1) and 16.7 per cent (G3) of the total score for that sound respectively. The lowest scores by G2 were found in /z/, with 13.9 per cent of the total score. The phoneme with the best scores in the pre-test was /ʌ/, also showing the highest performances in the post-test.

Table 36. Mean scores, standard deviations and degree of improvement in the imitation task

	/æ/			/ʌ/			/ə/			/z/		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	0.6 (0.9)	1.3 (1.5)	0.7	1.2 (1)	2.2 (1.3)	1	1.3 (1.2)	1.4 (1.2)	0	0.8 (1.2)	1.2	0.3
%	15.4	33.7	18.3	30.8	55.8	25	33.7	34.6	1	21.2	29.8	8.7
G2	0.9 (1.2)	1.0 (1.1)	0.1	2.0 (1.2)	2.5 (1.6)	0.6	1.0 (1.1)	1.2 (1.4)	0.2	0.6 (0.8)	0.9	0.3
%	22.2	24.1	1.9	49.1	63	13.9	25.9	30.6	4.6	13.9	22.2	8.3
G3	0.7 (0.9)	1.3 (1.3)	0.7	2.1 (1.5)	2.8 (2)	0.7	1.1 (1.5)	1.1 (1.3)	0	0.9 (1.2)	1.1	0.2
%	16.7	33.3	16.7	52.8	69.4	16.7	27.8	27.8	0	22.2	27.8	5.6

The results from the sentence-reading task reveal that learners in the experimental groups made significantly larger improvements than learners in G2. An ANOVA revealed a significant effect of the time variable ($F(1,59) = 44.7$, $p = 9.1e\text{-}09 < 0.05$) and an interaction effect between group and time ($F(2,59) = 3.97$, $p = 0.0241 < 0.05$). However, even though G3 was the group showing the most substantial improvements (Figure 65), Holm post-hoc tests revealed that the differences were only significant for G1. The improvement made by G3 was of 6.1 points (10.5%), followed by G1 with an improvement of 5 points (8.6%) and, finally, G2 with 2.1 points (3.6%).

As is the case in the other tasks, the learners' post-test scores show that they were all below 35% of the maximum scores for this task after receiving instruction. Out of the 58 points that could be obtained in this test, participants in G1 finished the study with 15.3 (26.3%), participants in G2 with 10.5 (18.1%) and participants in G3 with 18 (31%).

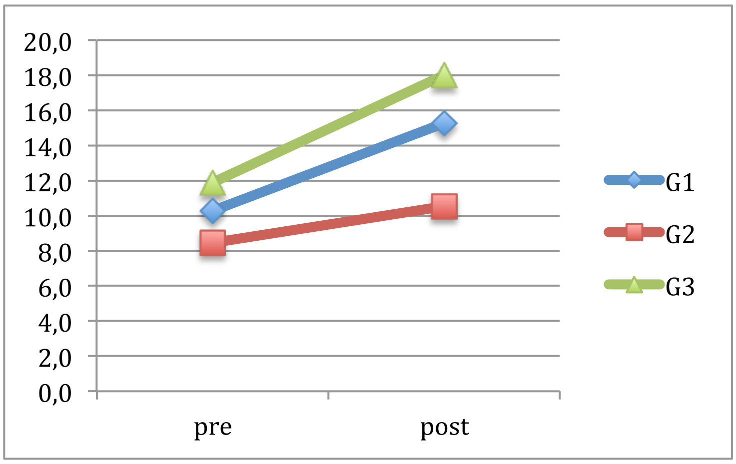


Figure 65. Average scores and degree of improvement for the sentence-reading task

Focusing on the differences among sounds, the ANOVAs revealed a significant effect of the time variable for every sound (Table 37), but no interaction effects were found between time and group. That is, there were significant improvements from pre- to post-test, but the degree of improvement participants made was similar across groups.

Table 37. p-values for the time variable in the ANOVAs for each sound in the sentence-reading task

Sound	time variable	interaction between time and group
/æ/	$F(1,59) = 28.3$, $p = 1.66e\text{-}06 < 0.05$*	$F(1,59) = 2.99$, $p = 0.057 > 0.05$
/ʌ/	$F(1,59) = 17.1$, $p = 0.000113 < 0.05$*	$F(1,59) = 1.74$, $p = 0.182 > 0.05$
/ə/	$F(1,59) = 20.7$, $p = 0.0000274 < 0.05$*	$F(1,59) = 2.65$, $p = 0.078 > 0.05$
/z/	$F(1,59) = 9.31$, $p = 0.00341 < 0.05$*	$F(1,59) = 1.05$, $p = 0.357 > 0.05$

The data show that, overall, G1 and G3 obtained better results than G2, except for /ʌ/, where improvements by G3 were only slightly higher than those by G2. The improvement made by G1 was similar across sounds, while G3 obtained a greater improvement for /æ/ than for any other sound. Table 38 presents the mean scores for each sound at pre- and post-tests, the maximum score for each sound (Max), their standard deviations and the degree of improvement (imp).

Table 38. Mean scores and improvement made in the sentence reading task for each sound

Total scores sentence-reading task												
	/æ/			/ʌ/			/ə/			/z/		
Max	14			14			15			15		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	1.3 (1.5)	2.9 (2.8)	1.6	4.4 (2)	5.8 (2.9)	1.4	3.2 (2.8)	4.3 (3.1)	1.2	1.3 (3.3)	2.2 (4)	0.8
%	9.6	20.9	11.3	31.6	41.8	10.2	21	29	7.9	9	14.4	5.4
G2	1.4 (2.4)	2.2 (3.3)	0.8	3.9 (2)	4.5 (2.1)	0.6	2.5 (2.8)	2.9 (2.8)	0.4	0.6 (0.9)	0.9 (1.6)	0.3
%	10.1	15.9	5.8	28	32.3	4.2	16.5	19	2.5	4.2	6.2	2
G3	1.2 (1.8)	4.2 (4.1)	3	4.3 (2.6)	5.2 (3.1)	0.9	4.4 (4.7)	5.6 (4.9)	1.1	1.9 (3.3)	3 (5.7)	1.1
%	8.7	30.2	21.4	31	37.3	6.3	29.6	37	7.4	12.6	20	7.4

In line with the results obtained in the imitation task, some of the standard deviations encountered in the sentence-reading task are exceedingly high, sometimes even higher than the mean. This shows that the degree of improvement made by each group varied substantially across participants. That is, there may have been participants who benefitted considerably from instruction, while others may have remained stable or with very moderate improvements. As a case in point, two participants in G1 improved more than six points in one of the sounds (around 40% of improvement),[101] nine improved more than four points (around 26.6%), 11 improved more than two points (13.3%), and the rest were between two to 0 points or even had negative scores. A similar pattern was found in G3, with a participant improving up to 10 points in one of the sounds (66.6%), four participants four or more points (26.6%), and the majority between two and 0. As for G2, only two participants made an improvement over four points and only seven improved two or more points. The data for individual participants in this task can be found in Appendix 22.

As explained above, a number of novel words were included in this task in order to measure possible generalisation gains. In line with the results obtained from the total scores

[101] The maximum scores in this task were not the same for every sound, as there were two words that were omitted from the analysis for /æ/ and /ʌ/. Thus, the maximum scores for these two sounds was 14, and the maximum score for /ə/ and /z/ was 15. For the sake of simplicity, the percentages in this paragraph are calculated over a maximum score of 15 points.

in this task, significant effects of the time variable were found in all the sounds, although interactions between time and group only occurred for /ə/ (Table 39). In other words, with the exception of schwa, all groups made significant improvements in novel stimuli from pre- to post-test. As in the other tasks, the results from the Holm post-hoc test revealed that the differences in the degree of improvement made for schwa were only significant for G1.

With regard to familiar items, significant effects were found in the time variable for all sounds, but there were no interaction effects between the time and group variables for any sound. That is, the scores for familiar stimuli developed similarly for participants in the three groups.

Table 39. Resuls from the ANOVAs comparing all groups' performance for familiar and novel stimuli

	Familiar	
Sound	**Time variable**	**Interaction between time and group**
/æ/	$F(1,59) = 21.4, p = 0.0000206 < 0.05$*	$F(2,59) = 2.998, p = 0.0575 > 0.05$
/ʌ/	$F(1,59) = 12.6, p = 0.000761 < 0.05$*	$F(2,59) = 0.657, p = 0.5216 > 0.05$
/ə/	$F(1,59) = 15.4, p = 0.00023 < 0.05$*	$F(2,59) = 0.867, p = 0.4253 > 0.05$
/z/	$F(1,59) = 9.29, p = 0.00345 < 0.05$*	$F(2,59) = 2.91, p = 0.0622 > 0.05$
	Novel	
Sound	**Time variable**	**Interaction between time and group**
/æ/	$F(1,59) = 15, p = 0.000272 < 0.05$*	$F(2,59) = 2.997, p = 0.0575 > 0.05$
/ʌ/	$F(1,59) = 6.87, p = 0.0111 < 0.05$*	$F(2,59) = 1.841, p = 0.1676 > 0.05$
/ə/	$F(1,59) = 5.37, p = 0.024 < 0.05$*	$F(2,59) = 6.18, p = 0.00366 < 0.05$*
/z/	$F(1,59) = 4.87, p = 0.0313 < 0.05$*	$F(2,59) = 0.027, p = 0.9724 > 0.05$

The mean scores for each sound, including the percentages considering the maximum score for the task per sound are presented in Table 40. The data show that the experimental groups made more noticeable gaims than G2 for /ʌ/ and /ə/, but not for /æ/ or /z/. G3 outperformed G2 for /æ/, with an improvement of 17.8 per cent from pre- to post-test as opposed to the 7.4 per cent by G2. However, the improvement made by G1 was smaller than that by G2 (2.3%).

Table 40. Mean scores, standard deviations and rate of improvement for familiar and novel items in the sentence-reading task

Familiar items												
	/æ/			/ʌ/			/ə/			/z/		
Max.	9			10			10			10		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	1.0 (1.3)	2.5 (2.5)	1.5	3.2 (1.6)	4.0 (2.2)	0.8	2.5 (2.2)	3.1 (2.5)	0.6	0.7 (2.0)	1.3 (2.6)	0.5
%	11.5	27.8	16.2	31.9	40.4	8.5	25.4	31.2	5.8	7.3	12.7	5.4
G2	1.1 (1.9)	1.5 (2.7)	0.4	2.8 (1.5)	3.2 (1.6)	0.4	1.7 (2.0)	2.1 (2.1)	0.4	0.3 (0.6)	0.3 (0.7)	0.0
%	11.9	16.9	4.9	27.8	32.2	4.4	16.7	20.7	4.1	3	3	0
G3	0.9 (1.2)	3.0 (3.2)	2.1	3.3 (2.0)	4.0 (2.7)	0.7	3.0 (3.1)	4.0 (3.3)	1.0	0.9 (1.5)	1.8 (3.1)	0.9
%	9.9	33.3	23.5	33.3	40	6.7	30	40	10	8.9	17.8	8.9

Novel items												
	/æ/			/ʌ/			/ə/			/z/		
Max.	5			4			5			5		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	0.3 (0.5)	0.4 (0.8)	0.1	1.2 (0.8)	1.8 (1.0)	0.6	0.6 (0.9)	1.2 (1.1)	0.6	0.6 (1.4)	0.9 (1.6)	0.3
%	6.2	8.5	2.3	30.8	45.2	14.4	12.3	24.6	12.3	12.3	17.7	5.4
G2	0.3 (0.7)	0.7 (1.1)	0.4	1.1 (0.8)	1.3 (0.8)	0.1	0.8 (1.1)	0.8 (1.1)	0.0	0.3 (0.6)	0.6 (1.3)	0.3
%	6.7	14.1	7.4	28.7	32.4	3.7	16.3	15.6	-0.7	6.7	12.6	5.9
G3	0.3 (0.7)	1.2 (1.3)	0.9	1.0 (0.9)	1.2 (1.2)	0.2	1.4 (1.7)	1.6 (1.7)	0.1	1.0 (1.8)	1.2 (2.6)	0.2
%	6.7	24.4	17.8	25	30.6	5.6	28.9	31.1	2.2	20	24.4	4.4

The last task was the timed picture-description. The results revealed a significant effect of the time variable ($F(1,59) = 21.9$, $p = 0.0000175 < 0.05$) and an interaction effect between time and group ($F(2,59) = 4.58$, $p = 0.0142 < 0.05$). The results form the Holm post-hoc analysis showed that the significant effect was found in the performance by G1. The total pre-test scores across sounds were 1.6 for G1 (11.5%), 1.15 for G2 (8.2%) and 2.44 for G3 (17.5). In the post-test, these increased to 3.5 for G1 (24.7), 1.6 for G2 (11.6%) and 3.4 for G3 (24.6%). Nevertheless, the average scores of experimental and control groups were all below 2.5 points in the pre-test and only increased up to 3.5 points in the post-test, which can be considered to be rather low given that the maximum score possible for this task was 14 points.

An analysis of the scores obtained for each sound reveals that all groups made significant improvements from pre- to post-test for all sounds except for /z/. However, the improvements made by G1 participants for /ə/ were significantly higher than those by the other groups (see Table 41).

Table 41. Results from the ANOVAs comparing all groups' performance at pre- and post-tests in the timed picture-description task for each sound

Sound	Time variable	Interaction between time and group
/æ/	$F(1,59) = 14.6, p = 0.000324 < 0.05$*	$F(2,59) = 1.101, p = 0.3392 > 0.05$
/ʌ/	$F(1,59) = 7.52, p = 0.00806 < 0.05$*	$F(2,59) = 2.19, p = 0.1205 > 0.05$
/ə/	$F(1,59) = 12.2, p = 0.000921 < 0.05$*	$F(2,59) = 4.74, p = 0.0124 < 0.05$*

It must be noted that stimuli in this task were all items that appeared in training. They were selected from the most frequently occurring words in the app in the trial run explained in section 5.2.2c. However, the pre-test scores in this task are generally lower than the scores in familiar stimuli in the sentence-reading task, although there are some exceptions, such as G1's scores for /æ/ or G3's scores for /z/, which are higher in the timed picture-description task. The results from this task must be interpreted with caution, as the maximum score for each target sound in this task was very low. The learners' spontaneous production was measured with only four words for /æ/ and /ə/, and with three words for /ʌ/ and /z/.

As shown in Table 42, the improvement made by G1 was generally higher than the improvement made by the other two groups, with the exception of /æ/, where G3 made a slightly greater improvement than that by G1. The improvements made by G1 were considered to be significant for /æ/, /ʌ/, and /ə/. However, these were only significantly different than those by the other two groups for /ə/, where G1 improved 20.2 per cent as compared to the 2.8 per cent by G2 or the 8.3 per cent by G3. Nonetheless, in light of the results obtained in the other tasks, the post-test scores for all groups were relatively low considering the total scores for the task. The highest post-test scores were 30.8 per cent for /ə/ by G1, or 37 per cent for /ʌ/ by G3 (see Table 42).

Table 42. Mean scores, standard deviations and degree of improvement for each sound in the timed picture-description task

	/æ/			/ʌ/			/ə/			/z/		
Max	4			3			4			3		
	pre	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
G1	0.69 (0.9)	1.2 (1.4)	0.5	0.31 (0.5)	0.7 (0.8)	0.4	0.42 (0.9)	1.2 (1.3)	0.8	0.19 (0.7)	0.3 (0.7)	0.1
%	17.3	29.8	12.5	10.3	24.4	14.1	10.6	30.8	20.2	6.4	10.3	3.8
G2	0.26 (0.6)	0.5 (0.8)	0.2	0.37 (0.6)	0.4 (0.7)	0.1	0.48 (0.9)	0.6 (0.8)	0.1	0.04 (0.2)	0.1 (0.4)	0.1
%	6.5	12	5.6	12.3	14.8	2.5	12.0	14.8	2.8	1.2	3.7	2.5
G3	0.33 (0.7)	0.9 (1.1)	0.6	0.9 (0.9)	1.1 (1.1)	0.2	0.7 (1.1)	1 (1.2)	0.3	0.5 (0.9)	0.4 (0.7)	-0.1
%	8.3	22.2	13.9	29.6	37.0	7.4	16.7	25	8.3	18.5	14.8	-3.7

Questionnaires

As for the learners' responses to the questionnaires, 59 participants completed the initial questionnaire and 49 completed the final one. Since all participants acted as experimental group and received the same type of instruction, no distinctions are made among groups when discussing the results. The questionnaires for study 2 can be found in Appendices 23 and 24.

Of the 59 participants who completed the questionnaire, 58 owned a smartphone (98.3%). Regarding the time learners use their mobile phones daily, 37 students (62.7%) claim to use it more than two hours, 15 (25.4%) between one and two hours, four (6.8%) between 30 minutes and an hour, two (3.4%) between 15 and 30 minutes, and one (1.7%) less than 15 minutes. When asked about the daily amount of time they would be willing to spend using a language learning app, one participant said that 'more than one hour a day' (1.7%), 16 said 'between 30 minutes and an hour' (27.1%), 40 said 'between 15 and 30 minutes' (67.8%) and two 'less than 15 minutes a day' (3.4).

The learners' responses to the initial questionnaire reflect their willingness to use language learning apps, with a mean rating of 4.3 (1). The perceived potential of mobile apps for language learning was 4.2 (0.9). At the beginning of the study, 24 participants (40.7%) had already used mobile applications for some kind of educational purpose and 35 (59.3) had not. The apps they mentioned in the open-ended questions included dictionary apps, automatic translators, apps to learn irregular verbs or practice phonetic transcription, or apps to learn about geography and general knowledge.

A number of 5-point Likert-type questions addressed the learners' opinion towards the app, including usability, design, the contents covered, the amount of practice or feedback provided, among others. These are presented in Table 43. The learners' responses reveal that, overall, they considered the app to be useful (MS = 4.4) and that they enjoyed using the app (MS = 3.8). Besides, most students considered the app's design to be appealing (MS = 4.1). Seventeen students claimed to have already used EFL Oxford books with similar phonemic symbols.

The contents addressed in the app were perceived as interesting (MS = 4) and relevant to the students' daily use of English (MS = 4.2). Besides, learners believed the app had helped them improve their perception (MS = 4.6) and production (MS = 4.3) of the target sounds.

Most students considered the app easy to use (MS = 4.8). Only one student claimed to have had some kind of technical problems when using the app. As reported by this participant, the app occasionally changed the option the student had marked when it was correct and it acted as though the student had chosen the wrong option. With regard to the level of difficulty in the activities, students did not perceived them to be difficult (MS = 1.7).

As for the amount of practice and feedback offered, the mean scores decreased slightly from the ones reported above when assessing the number of sample words included in the phonemic chart (MS = 3.6) or in the activities (MS = 3.4). However, feedback was regarded to be adequate (MS = 4.1) and enough to let learners know why their answers were wrong when they made mistakes (MS = 4.2).

Although the score was also slightly lower for the item addressing learners' willingness to continue using the app (MS = 3.8), the majority claimed that they would be willing to keep using it if new levels and activities are released (MS = 4.4). Finally, the mean scores for the last two items suggest that even though the MS for the item asking whether the app was entertaining was not very high (MS = 3.7), the app was not considered to be excessively monotonous either (MS = 3.3).

Table 43. Mean scores for the participants' responses to the Likert-type items in the final questionnarires for study 2 including standard deviations in parentheses

	Item	MS
1.	Overall, I think the English File app is useful	4.4 (0.8)
2.	The contents covered in the app are relevant to my daily use of English	4.2 (0.8)
3.	The app has helped me perceive the target sounds better	4.6 (0.6)
4.	The app has helped me produce the target sounds better	4.3 (0.8)
5.	I find the aspects covered interesting	4 (0.7)
6.	I have enjoyed using the app	3.8 (0.9)
7.	I find the app's design appealing	4.1 (0.8)
8.	I find the app easy to use	4.8 (0.6)
9.	I have found activities difficult	1.7 (0.8)
10.	The number of sample words for each sound in the phonemic chart is enough	3.6 (1.2)
11.	The number of sample words in the activities is enough	3.4 (1.1)
12.	The feedback offered by the app was adequate	4.1 (1.2)
13.	When I chose the wrong answer, the feedback provided helped me know why my choice was incorrect	4.2 (1)
14.	I will keep using the app once the study is over	3.8 (1)
15.	I am willing to keep using the app if new levels and activities are released	4.4 (0.9)
16.	The app was entertaining. I did not get bored	3.7 (1.1)
17.	The app was monotonous	3.3 (1.3)

Furthermore, when asked whether they had seen the contents taught by the app somewhere else, 15 participants claimed to have seen them all during the phonetic course they were enrolled on and six emphasised the /s – z/ contrast in plurals. The rest did not answer this question.

Thirty-eight participants considered the duration of the study to be 'just the right amount of time', while five claimed that some more days would have been perfect in order to consolidate the target sounds, and five said it was a bit longer. One student said that (s)he would not have minded to carry on some more days if the app had been more dynamic. (S)he mentioned that the app was very monotonous.

When asked about the perceived appropriateness of the time they were required to use the app daily, 45 students (91.8%) said it was appropriate and four (8.2%) mentioned that it was excessive. In one of the open-ended questions, learners asked about the amount of time per day they considered necessary in order to acquire the target sounds with the app. Their responses ranged from 10 minutes to an hour, but they have been grouped into four broad categories for the sake of clarity. Three students (6.1%) considered they needed between 45 minutes and an hour, three students (6.1%) felt they needed between 30 and 45 minutes, 32 students (65.3%) mentioned times between 15 and 30 minutes, and 11 students (22.4%) mentioned 15 or less than 15 minutes. Thus, the vast majority considered the 20 minutes recommended by the researcher appropriate. As noted above, Twenty minutes were only an orientative measure, the amount of time they had to spend using the app was measured by the number of games they completed.

At the beginning of instruction, participants were encouraged to use the app in order to practise with sounds other than the targets recommended by the researcher, although they were asked to pay special attention to the latter. One of the open-ended items in the questionnaire asked them about further sounds they had explored. Twenty-four students (49%) said that they had concentrated exclusively on the target sounds proposed by the researcher. However, the rest of students mentioned explored other sounds and sound contrasts. The /f – v/ contrast was mentioned by two participants, as well as /dʒ/, the /ɒ – ɔː/ contrast was mentioned by one student, and /ʒ/ and /ʃ/ were mentioned by one student each. Surprisingly, two students mentioned the /s – z/ contrast as something extra, even though these were considered to be among the target sounds in the study.

Regarding the questionnaire item dealing with students' attention to the feedback received at the end of each level, twenty-eight (57.1%) said that they 'always' paid attention to the feedback, 15 (30.6%) said that they 'often' did it, and six (12.2%) said that they 'sometimes' did it.

When asked to select the activity of the app they perceived to be most useful, 16 (32.7%) selected the phonemic chart, 27 (55.1%) chose activity 1 (with audio) and six (12.2%) chose activity 2 (without audio).[102] However, when asked to rate the perceived usefulness of each activity, activity 1 received a mean score of 4.8 (0.4), followed by activity 2 (MS = 4.4; 0.8), and finally, the phonemic chart (MS = 3.9; 0.9).

One of the open-ended items asked students about alternative ways in which they would have used the app had the researcher not recommended this order (i.e. phonemic chart, activity 1, activity 2). Fourteen students (28.6%) said that they would have followed the same order. However, five (10.2%) said that they would have done the activity without sound first, and then the activity with sound – one pointed out that this would allow them to realise what they do not know and in order to concentrate on this when using activity 1. Five students (10.2%) agreed that they would have used the phonemic chart less, arguing that they preferred to go straight to the activities. Two students (4.1%) mentioned that they would alternate the order of activities, one recommended using activity 1 one day and activity 2 the next, whereas the other recommended starting with a different activity each

[102] Even though the phonemic chart is not an activity per se, it is included here given that it offers learners the opportunity to hear the target sounds in various phonetic contexts and under different spellings.

day. One student (2%) mentioned that (s)he would prefer to focus simply on activity 1, while another one claimed to prefer activity 2. A number of participants seem to have misinterpreted this question, as they simply replied 'yes' despite the fact the question directly asked them 'If the researcher had not given any guidelines as to how to use the app, how would you have used it?' Finally, the responses by three participants suggest that they either misunderstood the question or the instructions at the beginning of the study, as their recommendation for change was to follow exactly the same order proposed by the researcher.

Eight students (16.3%) claimed that they 'always' repeated words out loud as they were played in the app, 31 (63.3%) said that they did it 'almost always', and 10 (20.4%) claimed that they did it 'sometimes'. However, only nine students (18.4) used the app's recording function.

As regards the stimuli featuring the target sound in the app, 47 participants claimed to be familiar with the vocabulary, one claimed not to be familiar with it, and one acknowledged that (s)he did not pay attention to the words. The mean score rating the usefulness of the vocabulary featured in the app was 3.8 (0.8). Twenty-two participants (44.9%) thought that the vocabulary was 'very' repetitive, 25 (51%) thought it was 'quite' repetitive, and 2 (4.1%) said that it was 'sometimes'.

As for the time of the day in which learners preferred to use the app, four students claimed that they preferred the morning, one added that this was the time (s)he devoted to studying, another one said that (s)he wanted to complete the games 'as soon as possible' in order to have the rest of the day free, and another one said that it was a way of establishing a routine and not forgetting about it. Six students mentioned the afternoon because this was a time at which they felt relaxed, before starting their evening of study. One added that (s)he preferred this time because she would be tired at night and would not be able to concentrate. Seven participants mentioned the evening, most of them said that it was the time of the day when they had most free time and felt more relaxed. One participant mentioned that she liked to use the app while waiting for the bus, claiming that it was a good way to entertain herself. Finally, 15 participants opted for the night. The reasons they gave were that they were more relaxed, that it was the only time of the day they were free, or that there was less noise and they could concentrate more easily on the sounds.

One of the open-ended items asked learners for recommendations for future implementations of this study. Sixteen students (32.7%) suggested including more sample words. Several students pointed out that so much repetition became monotonous, as sample words for each sound are often the same. One student said that it was even possible to learn the answer by heart without having to listen to the stimulus in order to mark the right option. Also, one student recommended focus on a different sound each day.

Regarding the amount of time learners had to use the app, four students (8.2%) recommended using the app some more days in order to automatise the sounds. However, two students suggested using the app some more time every day, while three recommended reducing the number of activities per day.

Students also made some suggestions in terms of how to improve the app. These included listening to the sound in sentences (mentioned by two students), adding an activity to practise transcription, or including intonation. One participant also mentioned that it would be interesting to be able to see how many games had been played as well as the scores received. Two participants said that it would be convenient to underline the syllable where the target sound is for activity 2. One mentioned that it was easy to get confused with sounds from other syllables, therefore leading to mistakes despite learners knowing the right option. Another participant suggested adding more options as distractors, since the identification with only two sounds was considered to be easy.

With regard to feedback, one student mentioned that it would be useful if words that had been marked as wrong during the activities could be played back at the end. In addition, one mentioned that it would be very helpful to see the transcription of words immediately after they made mistakes; (s)he said it was not very convenient to wait until the last of the 10 words in order to see the transcription of the word you got wrong. Another student claimed that it would be interesting to be able to know why their choices were incorrect in some cases (e.g. possibly receiving an explanation about the pronunciation of <-ed> endings). Finally, two emphasised that the app was very monotonous at times, as the routine was always the same and the words tended to be repeated. Some of the recommendations included adding more levels where vocabulary became increasingly more complex, for example.

When asked whether they would pay 5.5€ for this app, 13 students (26.5%) said that they would, although one pointed out that it was ‘a bit expensive’ for what it offered. The rest said that they would not buy this app, 14 (28.6) because it is too expensive and 21 (42.9%) because they never spend money on apps. Additionally, they were asked about the ideal price for this app. Seven students (14.3%) recommended between 1 and 2€, 13 students (26.5%) were between 2 and 3€, 11 (22.4%) were between 3 and 4€, and 9 (18.4%) between 4 and 5€. Thus, most participants think that a lower price would be suitable for this app.

5.4.4 Discussion

The results reported above offer empirical data to provide an informed answer to the study's research questions.

RQ1

The first research question was concerned with the effect of the app-based training on the learners' perception of /æ/, /ʌ/, /ɑː/, /ə/ and /z/. It was hypothesised that receiving training with the app would improve the learners' perception of the target sounds, as reflected by their scores in the discrimination and identification tasks.

The data from the discrimination task show that, in general, the improvement made by participants in the three groups was very moderate. Participants made significant improvements from pre- to post-test in their discrimination of the /ʌ – ɑː/ contrast and in the catch triads for /ʌ/and /z/. However, the improvements were similar for all groups, which indicates that instruction did not have a significant impact on the learners' discrimination of the target sounds. On the other hand, the results from the identification task reveal a significant effect of training. Participants in the experimental groups (G1 and G3) made substantial improvements from pre- to post-test and these were significantly different from those by the control group (G2). Moreover, when G2 acted as experimental, their performance also improved significantly as compared to their progress as control.

Focusing on the scores from the imitation task, the results isolating the total scores for each sound revealed that, although groups acting as experimental always made greater improvements than the control group, the differences among groups did not reach significance for all sounds. Likewise, not all groups benefitted from instruction in the same way. G1 made significantly greater improvements than G2 for /ʌ/, /ɑː/, /ə/, and /z/, whereas G3's improvement was only significantly different than the one by G2 for /ə/. Nevertheless, despite fact that G2 made considerable improvements when acting as experimental group (sometimes even higher than those by G1), these did not reach significance for any of the sounds compared to the improvements made when they acted as control. In any case, the data suggest that despite the fact that differences among groups did not reach significance for all sounds or for all groups, the instruction had a very positive effect on the learners' perception of the target sounds.

The frequency of appearance of each sound does not seem to have affected the improvements made for each sound. /ə/ was the sound that appeared most often, with FOC of 8.8 per cent. This was the only sound for which improvements reached significance for G3, which could be related to its high FOC. However, the improvements made by groups 1 and 2 (when acting as experimental) were higher for /ʌ/, even though the latter sound has a FOC of only 3.2 per cent. Likewise, /æ/ was the second sound that appeared most often, with a FOC of 6.8 per cent, but no group made significant improvements in the identification of this sound.

Studies training learners' perception with identification tasks often use HVPT paradigms where participants are presented with highly variable stimuli that help them direct their attention to phonetic cues in order to discern between similar target categories (e.g. Bradlow et al., 1997; Logan et al., 1991). Studies using highly variable stimuli have promoted significant improvements in the learners' perception (Rato, 2014; Uther et al., 2007) and production (Thomson, 2011; Wong, 2015) of problematic contrasts with as little as around 5 hours of training.

Nonetheless, unlike in HVPT paradigms, the stimuli in the present study were far from highly variable. Indeed, they were rather repetitive. Tokens were pronounced by a single speaker and they were limited to a few sample words for each sound. The approach adopted did not rely so much on learners' paying attention to phonetic cues by cross-sound comparison, but on fostering an adequate conceptualisation of the target sounds by exposing them to different instantiations of English phonemes and trying to familiarise them with their phonemic symbols. This was hypothesised to help them create mental categories for the target sounds, therefore allowing them to 'classify' future instantiation of the target sounds as belonging to these categories (see section 1.1.5). The results show that training exerted a significant effect on the learners' perception of four of the target sounds despite the limited variability of stimuli and the limited amount of training. Therefore, longer periods of instruction with a greater stimuli variation should yield much better results.

RQ2

The second research question was concerned with possible improvements in the learners' production of /ʌ/, /ɑː/, /ə/, and /z/ as a result of the perceptual training offered by the app. It was hypothesised that training with the app would not only help learners improve their perception of the target sounds, but also their production. The results from the total scores for each sound in the production tasks reveal that instruction had a significant effect on the pronunciation of /æ/, /ʌ/, and /ə/ for participants in G1. The improvements made by participants in G3 were superior to those by participants in G2 for all sounds, sometimes even higher than those by participants in G1. Nonetheless, the improvements made by G3 did not reach significance.

Furthermore, a closer look at the improvements made for each sound showed that improvements did not reached significance for these sounds in all tasks. As regards the imitation task, significant differences in the improvement made among groups were only found for G1 in their pronunciation of /æ/. There were also significant differences in the scores from pre- to post-tests for /ʌ/ and /z/, but these were similar across groups, which indicates that instruction only had an impact on the learners' imitation of /æ/. The learners' improvements in the sentence-reading task were very similar across groups. Although there were significant improvements from pre- to post-test for all groups, differences among groups did not reach significance. This indicates that instruction did not have a significant effect on the learners' controlled production of the target sounds. Finally, the results from the learners' performance in the timed picture-description task show that there was a significant effect of training in the learners' spontaneous production of the target sounds. A closer inspection of the data for each sound revealed that the differences among groups were only significant for /ə/ by participants in G1, whereas all groups improved similarly in their spontaneous production of /æ/ and /ʌ/.

The fact that there were significant improvements for /ə/ in G1's spontaneous production but not in their controlled production is rather surprising, given that performing adequately in the sentence-reading task is supposed to be easier for learners than in the timed picture-description. Common mispronunciations of this sound by Spanish EFL learners often result from orthographic interference, yielding examples of vowel substitution such as *[ˈfeimus], *[ˈfeimos], or *[ˈfeimous] for *famous*, *[ˈafrika] for *Africa*, *[ˈpiktur], *[ˈpiktʃa] for *picture*, or *[ˈdeindʒerous], *[ˈdeindʒeros] or *[ˈdeindʒerus] for *dangerous*.[103] Thus, given the natural tendency of English vowels to be reduced to schwa in unstressed positions in connected speech, it may be the case that the higher cognitive

[103] These examples of interlanguage were obtained from the production data in this study, which are in line with the data collected by Monroy (2001).

demands imposed by the task prevented learners from focusing on orthography, and therefore, avoided interference.

Focusing on the learners' most 'basic' ability to produce sounds, that is, whether they can attain the sounds in production, the results from the imitation task in the post-test show that their mean scores were rather low as compared to the maximum scores for the task. The highest scores were found for /ʌ/ (55.8%-69.4% range). However, the learners' scores for /æ/, /ə/ and /z/ were extremely low, which suggests that the learners' phonetic categories for these sounds were not sufficiently developed in terms of production. Nonetheless, the scores obtained in the discrimination and the identification task show that participants were much closer to native-like proficiency in terms of perception (with scores always higher than 50%, and most of them over 60 or 70%). As the data from study 1, this adds to the evidence supporting the precedence of perception over production.

The low scores obtained in the sentence-reading task and in the timed picture-description task suggest that despite the learners' relative capacity to discriminate and identify sounds, they are still unable to articulate sounds correctly all the time (as evinced in the imitation task), and they have even more problems in tasks that require them to be familiar with the phonological composition of words (i.e. the sentence-reading task and the timed picture-description task).

It would be interesting to compare the learners' performance in each of the production tasks in order to explore differences resulting from the demands imposed by each task. However, no reliable comparisons can be established given that the number of stimuli in each task were not the same (four novel stimuli in the imitation task, 10 familiar and five novel in the sentence-reading task, and four familiar items in the timed picture-description task). Furtheremore, care should be taken to control possible effects of spelling on the learners' pronunciation, as well as lexical familiarity. An anecdotal observation comparing the learners scores from each task (considering only novel stimuli in the sentence-reading task) shows that, in line with the results obtained in study 1, the learners' scores were generally highest in the imitation task, followed by the sentence-reading task, and finally, the timed picture-description task.

RQ3

The third research question addressed whether gains obtained in the perception and production of the target items could be transferred to novel lexical items. It was hypothesised that learners receiving instruction with the app would not only improve their perception and production of the target sounds in words that appear in training, but also of novel words. The results from the identification task revealed significant differences in the improvement made by participants in G1 for novel stimuli as compared to the improvements made by the other two groups. However, these differences were not significantly different from those by the other groups when considering the scores for individual sounds. With regard to production, the differences among groups for novel stimuli only reached significance for G1's progress with /ə/, the only sound showing significant improvements in the learner's spontaneous production in the timed picture-description task.

It is important to note that in phonological acquisition research, generalisation is often tested by exploring possible transfer of the benefits obtained under one training condition to another condition, such as novel speakers, novel phonetic contexts, etc. (see e.g. Carlet & Cebrián, 2014; Rato, 2014; Thomson, 2011). As explained in section 3.1.1, testing stimuli in this study and in study 1 were always 'novel' in terms of speaker voices, given that learners were never exposed to the voices employed for testing during training. Thus, whenever improvements reached significance in the perception tasks in these two studies, this can be interpreted as generalisation to novel speakers.

Given that the training materials were not designed by the researcher, generalisation to different phonetic contexts was not controlled for. Learners were exposed to the variability offered in the EFP app, the same degree of exposure anyone would have when using the app on their own. However, testing stimuli from the discrimination task and the imitation task did include a range of phonetic environments in order to measure the learners' perception and production regardless of the context (see section 5.2.2d).

Generalisation to novel words in the sentence-reading task implies that learners are not only able to articulate certain sound correctly, but also know where they need to use the sound. This is the most difficult type of generalisation, given that learners need to become aware of the phonological composition of every word, especially difficult in English due to the lack of a one-to-one correspondence between phonemes and graphemes. Participants in this study generalised their improvements to novel lexical items for /ə/, but not for /æ/, /ʌ/, or /z/. This suggests that they may have become aware of the fact that vowels are commonly reduced to schwa in unstressed syllables. However, pronouncing /æ/, /ʌ/, or /z/ adequately in this task would require participants to be familiar with the phonological make-up of words featuring these sounds, which is more difficult given their different orthographic representations. Even though improvements of this kind cannot be discarded with longer periods of instruction, these have not taken place in this study, probably due to the rather limited variety in the training stimuli.

As a case in point, a comparison of the mean post-test scores for familiar and novel stimuli reveals that scores were generally higher for familiar than for novel items. It was expected that learners would obtain better scores for lexical items that had appeared in training, as they should be familiar with their phonological make-up. However, the fact that /ʌ/ and /ə/ were the sounds with the highest scores in both familiar and novel items, while /æ/ and /z/ showed comparatively much lower scores suggests that this may be due to the influence of spelling.

The low scores for /æ/ and /z/ in novel stimuli could be explained by the fact that the former is represented by the letter <a> in spelling, a common grapheme representing also other sounds. Likewise, familiar items for the latter sound always featured /z/ with spellings such as <s> (e.g. *music*) or <x> (e.g. *exams*), graphemes that coincide with the common spellings for /s/. This may have led learners to simply use /s/, the sound they already know and which is also represented by the grapheme <s>. This finds further support in the fact that the learners' scores for /z/ in novel items were considerably higher than their scores for familiar items, as the former were always represented by <z> in orthography (e.g. *zebra*) and may have given learners the clue that it was /z/ the sound they required. Nonetheless, with regard to /ʌ/, the sound with the best scores in novel items, learners may have associated the spellings featured in the app (<u> as in *hundred*, <o> as in *son*, or <ou> as in *young*) with the ones in novel words (<u> in *gum* and <o> in *brother*).

Finally, the considerably higher improvement made for schwa in novel items could also be explained by the fact that it was featured in 22 different items in training, as compared to the 10 items for /ɑː/ and /ʌ/ or the nine for /s/ and /z/. This may have helped learners became aware of a wider range of possible spellings for this sound.

RQ4

The forth research question addressed the impact of the app-based instruction on a group of average language learners. It was hypothesised that training with the app will foster similar improvements in English majors taking a phonetics course than on learners not receiving training in phonetics. The results show that the improvements made by G3 only reached significance as compared to the control group on their overall identification scores and on their identification of /ə/. However, the average improvement made by participants in G3 was similar or even superior to the one made by G1 in almost every task. The fact that G1's scores reached significance for many more sounds may be due to the low number of participants in G3 as compared to the other two groups. Although the overall mean score of G3 may have increased considerably in some tasks due to several good participants, the increase did not reach significance considering the group as a whole. This indicates that despite the fact that there were some learners who benefitted very little from instruction, the approach had a positive effect on the pronunciation of some learners. Even though the improvements made by G3 were not significant in some tasks, the results are encouraging given that the target phonemes tend to be fossilised in the interlanguage of very advanced FL learners.

RQ5

Finally, the last research question addressed the learners' perceptions towards the app-based instruction. The learners' responses to the questionnaires offer qualitative and quantitative data in order to answer this research question, but they also provide interesting insights into aspects that have to be considered when implementing this type of instruction, such as the learners' e-routine, familiarity with the technology, etc.

The participants' responses indicate that all students but one had their own smartphones and that the majority used them between 15 and 30 minutes a day. The only student who did not have a smartphone used her mother's phone instead. This may have limited the convenience of using the app for this individual participant, as she may not have had access to the phone whenever she needed it. However, this did not affect her completing the daily games.

The amount of time learners use their phones daily and the high mean scores in the items canvassing their willingness to use language learning apps suggest that implementing this type of instruction in their routine should not be problematic. The vast majority of students considered the amount of time they were asked to use the app daily appropriate. In fact, when asked to recommend an ideal amount of time in one of the open-ended questions, most students said between 15 and 30 minutes – although there were some who acknowledged they would need some more time in order to master the target sounds.

The learners' answers to the Likert-type questions suggest that they considered the app to be useful and that it helped them improve their perception and production of the target sounds. However, their responses regarding their perceived usefulness of the different activities were contradictory. The majority chose activity 1 as their preferred activity, followed by the phonemic chart, and finally, activity 2. Nonetheless, when asked to rate the usefulness of each of them, activity 1 received the highest score, followed by activity 2, and finally, the phonemic chart.

Limitations and directions for future research

The present study presents a number of limitations that must be acknowledged, some of them due to the app employed, others due to the design of the study itself. One of the learners' major concerns with the app was the limited amount of practice offered, as shown in the mean responses to the Likert-type questions or in the comments to the open-ended questions. Most of the recommendations students gave for improving the app were including more words and making the app less monotonous. In fact, despite their positive reactions to the instruction, very few claimed to be willing to purchase the app themselves. This is one of the major limitations of the current study, as instruction was not designed by the researcher but had to be adapted to already-existing materials.

The app selected for training included a very limited amount of sample words for each target sound. In the trial run carried out in order to explore the app's stimuli, there were 12 words for /æ/, 10 for /ɑː/ and /ʌ/, nine for /s/ and /z/, and 22 for /ə/. This is problematic given that learners may end up doing the activity correctly because they have learnt the words by heart, and not because they are identifying the target sound, as pointed out by one of the participants. This is not considered to be an issue in the present study given that significant differences not only occurred for familiar items, but also for novel words. However, as pointed out above, offering a wider range of sample words would be paramount in order to help learners generalise improvements to other words. As shown in the production data from study 1 and from this study, even if learners are able to pronounce sounds correctly, one of the most difficult things for FL learners is to know when to pronounce those sounds (i.e. in what words and for what spellings).

A further limitation is that the app does not allow users to choose the sounds with which they want to practise. Instead, learners are presented with a randomised set of words featuring different English sounds and the target sounds only appear ocassionally (see section 5.2.2c). Even though this may be helpful for learners to become familiarised with the phonemic inventory of English, it limits the amount of exposure to the target sounds considerably. Other apps, like *Sounds* (Macmillan Publishers Ltd., 2011), offer users the opportunity to choose the sounds they want to practise with, the number of trials they want to have, as well as the possibility to buy new word lists dealing with a range of topics.

This would have been very convenient for this study, although as noted above, the app was not compatible with all smartphone models.

Moreover, several participants mentioned that they would improve the feedback offered by the app. Even though users were indicated whether their answers were right or wrong with a tick or a cross immediately after they made their choices, they only saw the transcription of the words showing the right answer at the progress screen after every 10 games. Besides, in this app, the feedback provided after the responses to individual stimuli disappeared immediately, thus leaving little time for learners to assimilate why their response was wrong. A convenient alternative would be to either show learners the correct transcription after each mistake, as suggested by some participants, or to make them listen to the stimulus again in order to help them associate the sound with the correct answer (see e.g. Rato, 2014).

Additionally, it is important to note that learners in this study were allowed to use the app at any time. This is considered to be positive since it reflects the way learners use the app autonomously. However, the fact that learners are able to choose when to use the app may affect their performance (and consequently the results) if they use the app at times when they are too tired. As a case in point, some learners completed the activities very late at night (ranging from 23 pm to 2 am). The time at which learners completed the activities was not controlled for in this study. Even though the time at which participants took the screenshots can be checked in their Dropbox uploads, it is extremely difficult to control this manually. Using an app designed by the researcher would have allowed a much stricter control of the learners' performance, including the exact time spent on each task, number of mistakes across time, etc. Uther et al. (2007) did not find a relationship between usage time and improvements in performance, although they claim that their data were quite homogeneous. Hence, future research should investigate possible effects of the time at which learners use the app or the amount of time spent on each task.

Furthemore, the app used in this study did not evaluate the learners' production in any way. Offering learners' feedback on their pronunciation is extremely difficult due to the limitations of current ASR technology (see section 1.3.2b). Thus, the approach adopted here attempted to foster improvements in the learners' perception under the assumption that these would also improve their production (see section 1.1.5). Nonetheless, some kind of feedback of the learners' attempts at pronouncing the target sounds would have been beneficial for students. Even though state-of-the-art ASR technology is far from perfect, recent studies have shown that learners can improve their articulation of FL sounds through mobile-based practice using speech-to-text recognition applications (Liakin et al., 2014).

One of the recommendations learners made was to make the app less monotonous, as the EFP app always follows the same identification pattern. Some existing commercial apps include different gamification techniques which make using the app much more entertaining. For example, *Sounds* (Macmillan Publishers Ltd., 2011) offers the possibility to use a timer in the activities or lose 'lives' when making mistakes. Similarly, *Clear Speech* (CUP, 2011) offered segmental and suprasegmental practice through a range of game-like activities, such as a water blob that has to be matched to holes in a wall illustrating the stress pattern learners are hearing, or a stream of water which is either open or closed depending on whether sounds are fricatives or stops (see Fouz-González, 2012).

Some of the obstacles encountered in this study could have been overcome if the researcher had been able to use a self-developed app. One of the most important advantages of using one's own app is the possibility to measure the exact amount of practice learners receive on a sound. This presents great possibilities for speech acquisition research, as it would be possible to measure the exact amount of exposure learners need in order to attain a particular contrast or sound, for example. Also, with the necessary algorithms, learners could be offered more practice on the specific areas in which they make most mistakes. Uther et al. (2007) found no particular benefits from using this kind of adaptive training as compared to regular training over a period of two weeks. However, this may be especially useful for learners who aim to use the app autonomously for longer periods of time.

Another limitation of the present study is the reduced number of participants in G3. This group was set up in order to evaluate the effect of instruction on a group of average EFL learners, not receiving instruction in phonetics. Nonetheless, given the limited number of volunteers from the different degrees, this led to a group of few students with very different levels. Research has shown that training with HVPT techniques can be beneficial for both high and low proficiency learners (Wong, 2015). However, given the importance of phonetic symbols in the present study, it would be interesting for future studies to address whether this type of symbol-based training can help learners with no knowledge of phonetics with larger groups of students.

Additionally, as pointed out above, the length of the study was rather limited in order to foster substantial improvements in the learners' perception and production of the target sounds. Learners were asked to use the app for around 20 minutes a day over a period of ten days. Taken together, this adds up to a total of 3.3 hours, which is similar to the total amount of training in study 1. Both studies fostered significant improvements in the learners' pronunciation of some of the target sounds. However, the pre- and post-test scores are typically far from the maximum score possible.

A further limitation of the current study is that the learners' level of engagement in the study was not controlled for. As a case in point, the results from study 3 show a positive correlation between the participants' level of engagement and rate of improvement (see chapter VI). In this regard, some participants in study 2 claimed that they completed the activities early in the morning in order to finish them as soon as possible and have the rest of the day free. Some other participants completed them really late at night. Since learners were awarded points in exchange for their participation in the study, some may have completed the activities simply to get the reward, without actually trying to benefit from instruction. This could explain why some participants made such substantial improvements and improved only moderately. This suggests that despite the fact that not every participant benefitted in the same way from training, app-based pronunciation instruction can foster very good results for some participants.

Chapter VI. Exploring the potential of social networking services for pronunciation training: Twitter.

This chapter addresses the potential of social networking services (SNSs) to help learners improve their pronunciation of a number of English words commonly mispronounced by FL learners. More specifically, the chapter focuses on Twitter, an SNS with which users can interact in posts of 140 characters. Motivated by the various Twitter accounts that post 'words of the day' to teach learners new vocabulary items, or accounts that tweet on the pronunciation of various English words and sentences (see section 1.3.3), two studies were conducted recently in order to test Twitter's potential to help learners improve their pronunciation of commonly mispronounced words (Fouz-González & Mompean, 2012; Mompean & Fouz-González, in press).

The results obtained in those two studies suggest that Twitter is potentially very beneficial for pronunciation training. However, the results with a small subset of students with a different profile from a pilot study were not so positive (Fouz-González & Mompean, 2012) due to a lack of interest on the part of participants given their high exposure to English as part of their degree (see section 1.3.3). Thus, this study seeks to weigh the potential of this tool with a group of average learners of English as an FL. Like study 2, study 3 explores the way learners use this technology outside classroom settings.

6.1 Study 3

6.1.1 Aims, research questions and hypotheses

Study 3 intends to shed light on Twitter's potential to improve the learners' pronunciation of English words with a group of average language learners. As noted above, research has shown that Twitter can be very useful for this purpose, especially with highly motivated learners. The aim of this study was to test Twitter's potential with a group of regular language learners as something supplementary to an ESP course at university. The addresses the last two specific objective in this dissertation, namely we explore:

9. The potential of Twitter to improve Spanish learners' pronunciation of a number of problematic aspects of English often mispronounced by EFL learners due to sound-spelling correspondences or different stress patterns in cognates.
10. The learners' perceptions towards using Twitter for pronunciation training.

In order to address thess specific objectives study 3 presents three research questions and two hypotheses based on Twitter's potential as a teaching/learning tool (sections 1.3.3 and 2.4.3) and on the importance of explicit instruction and input enhancement for FL pronunciation training (section 1.1.5).

RQ1: Can Twitter help learners' improve their pronunciation of lexical items that are commonly mispronounced due to sound-spelling correspondences or different stress patterns in cognates?

H1: Sending learners a daily tweet with explicit explanations about the target aspects and audio/video illustrations will help them improve their pronunciation of the target words.

RQ2: Is Twitter-based instruction as effective when implemented with participants who volunteer to participate than with students who are offered a reward in exchange for their participation?

H2: Training with a group of students enrolled on an ESP course and rewarded for their participation will yield similar results to the ones obtained with a group of volunteer FL learners.

RQ3: What are the students' reactions towards using Twitter for pronunciation training?

6.2.2 Method

a) Participants

Participants in this study were recruited from the first year of a degree in Medicine at the University of Murcia during the first (autumn) semester of the academic year 2013/2014.

The researcher was teaching an *English for Medicine* course in that degree and had access to 260 students enrolled in the course. The study was advertised to all students, offering them 0.3 points of their final mark in exchange for their participation. One hundred and sixty-six students showed their interest to participate in this study. However, only 133 completed the initial questionnaire and only 121 completed the pre-test, followed training and completed the post-test. Thus, 121 students are considered to have participated in this study. They were 38 male, 83 female (mean age 19.1; SD = 3.7).[104]

At the beginning of the study, participants were randomly assigned to one of two groups: control and experimental. There were 42 students in group 1, acting as control group, and 79 students in group 2, acting as experimental group (N = 121).[105] As explained in section 3.1.3, participants were told that the researcher was investigating word-colour synaesthetic associations in FL learners, but that while they helped the researcher in his investigations, they would receive daily tips about different aspects of English that would complement their ESP course. Hence, participants in the experimental group received a tweet of the day on the problematic pronunciation targets and participants in the control group received a tip of the day covering aspects of English on a range of topics other than pronunciation (false friends, grammar tips, or clarifications about the contents in their English for Medicine course).

b) Training procedure

Training consisted in sending learners daily tweets with concise explicit explanations about the problematic aspects explained in section 2.4.3. Training took place over a period of five weeks, sending one new tweet (i.e. 'the tip of the day') from Monday to Friday. There were a total of 22 tweets. Tweets were sent around 10:00 in the morning so that they were available to learners for the rest of the day.

Since learners' mispronunciation of the target items was considered to stem from a lack attention to form, it was hypothesised that these types of errors could be corrected through form-focused instruction that helped learners 'notice' (Schmidt, 1990, 2010) features in the input they receive. The approach adopted in this study included input

---

104 The standard deviation is quite high due to the fact that there were three participants older than 30 years old. They had started the degree some years ago but they had not passed this ESP course yet for different reasons. The mean age without taking these four participants into account is 18.6 (SD = 1.7).

105 The control group had a comparatively small number of participants with respect to the experimental group, with almost half of participants. This is due to the fact that of the 166 participants who started the study, 40 dropped out before the study finished (see section 6.2.2a). Groups in the original distribution consisted of 78 participants in group 1 (control) and 88 participants in group 2 (experimental).

enhancement (Sharwood-Smith, 1993) and explicit instruction. Among the most common strategies for input enhancement are modifying typography to make certain features more salient (Han, Park, & Combs, 2008; Saito, 2013), or providing learners with metalinguistic information about the target features (Ioup, 1995). This study used these two strategies. Regarding the former, capital letters were used in order to mark stressed syllables in our tips and pronunciation-related familiar words were often used as examples alongside target words to promote meaningful learning (e.g. 'Listen to the 'ch' in *archives*. Careful! It's the same sound as the 'k' in *key* or the 'c' in *come*'). As for the latter, all tweets contain concise explicit explanations that draw the learners' attention to the target aspect. As in the podcasts study, directing the learners' attention to features that may otherwise go unnoticed was considered a way of enhancing the input even though audios were not altered in any way.

In the study by Mompean and Fouz-González (in press), participants were asked to confirm that they read our tweets by sending us a 'reading confirmation' tweet (henceforth RC), although they were informed that they could also include comments or questions about the tweets (those were classified as 'content tweets'). RC tweets were different from content tweets in that they former merely required learners to confirm reading by writing 'OK' or 'seen', whereas the latter reflected a deeper engagement on the part of learners.

RC and content tweets were a very suitable measure for the above-mentioned study, as learners were highly motivated to participate and were doing so in a completely voluntary manner, without any external rewards. In other words, if learners replied to a tweet claiming to have read it, there was not reason not to believe them, as they were not compelled to do so. However, given the substantial number of participants in this study and the external reward offered in exchange for their participation, RC tweets would not guarantee that participants actually read the information and followed the links provided to listen to the pronunciation model. They could simply send a tweet confirming reading even if they have not read it. Thus, this time participants were asked to reply to the researcher's tweets by writing down the sentence in which the target word was embedded in the audio/video file. As a case in point, for the tweet featuring the word *heir*, learners had to send a private message with the carrier sentence from the Smiths' video-clip the researcher shared (*I'm the son, and the <u>heir</u>, of a shyness that is criminally vulgar*). This was considered to be a more effective measure in order to control whether learners accessed the multimedia materials in the tweets, although it is not entirely reliable either.[106] Besides, it was a way of focusing their attention on the specific part of the file where the target word was and encouraging them to pay careful attention to it, as they had to spot it in the audio file and then write the whole sentence down.

A different account was created for each group ('@colourtweet1' for group 1; '@colourtweet2' for group 2). These two accounts were set up so that tweets were private to non-followers. That is, tweets were only visible to users that were approved by the researcher. Hence, users from group 1 were only allowed to see the control tweets and users from group 2 only the tweets addressing the target pronunciation aspects.

The accounts were called 'colour-tweet' because the synaesthesia pretence was maintained until the end of the study. The profile picture for each account was changed to a different colour every two or three days in an attempt to make learners believe that the researcher was really investigating possible word-colour synaesthetic associations with the target words (Figure 66). Students were told that the researcher was exploring whether the way instruction was conducted (e.g. changing the colour of the profile picture for different words) could have an impact on the colour the different words would evoke after having received those tweets with a particular colour. This was an attempt to draw their attention away from pronunciation being the goal of the study, reducing the possibility that they

[106] It must be noted that it is exceedingly difficult to find a way of ensuring that learners read the tweets and open the video/audio links. In the same way that some students cheat in exams, participants in a study of this kind might well send the carrier sentence to each other via private messages and simply copy and paste each other's answers without even reading the tweet. However, this is a limitation that cannot be overcome when using this type of tool not originally conceived for educational purposes. This issue may affect the results by those students in the experimental group who might have obtained poor scores in the post-test not because the Twitter-based instruction has not been effective, but because they have not read the tweets – perhaps simply pretending to participate in the study in order to obtain the 0.3 points.

would look words up in a dictionary or conscientiously study them for the post-test.

Figure 66. Sample tweet with the profile picture in green

In this study, participants were told to reply only via private messages in order to avoid participants in the control group seeing the target tweets and to facilitate keeping a register of participation.[107] In those cases where learners provided their own examples or asked questions that could be of interest to the whole group, the researcher copied them on the public timeline for that group so that other participants in the experimental group could see them too, therefore benefitting from vicarious learning. Furthermore, the researcher offered feedback on demand, including further clarifications, more examples, or even links to new audios/videos.

c) Training stimuli

Training stimuli consisted of 22 words featuring the target aspects explained in section 2.4.3. In previous studies (Fouz-González & Mompean, 2012; Mompean & Fouz-González, in press) training stimuli were selected after an analysis of pre-test productions. However, given the high number of participants in this study, a subset of 15 participants that were selected randomly for evaluation. The perfect set of target items would include items whose pronunciation had been problematic for 100% of participants, but this would have yielded a very low number of possible tweets, as not every participant had problems

[107] Users can interact in Twitter through private messages or public mentions. 'Mentioning' someone in Twitter means addressing this person publicly by typing an @ symbol before this person's username. When users interact through public mentions, all these interactions are shown in the user's timeline as posts that can be seen by all his/her followers.

with the same target words. Thus, items were included as part of the training if they had been mispronounced by 95% of participants. A higher rate of mispronunciation was considered to have led to an undesirably low number of tweets. Similarly, in an ideal situation, there would be the same number of items from each problematic aspect. Nevertheless, the final figures were dependent on the items with which most students had problems in the pre-test. Silent letters were featured in 11 tweets, grapheme-phoneme correspondences were featured in 10, and lexical stress in six.

A list with all the training stimuli, including tweets and links to the audio/video files in chronological order can be found in Appendix 25. Since this group of learners were not following a specific model of English in their instruction, and given the high number of participants (possibly with different preferences), phonological values for grapheme-phoneme correspondences as well as lexical stress in target items apply to standard BrE and AmE (see Jones et al., 2011; Wells, 2008). This was taken into consideration in the selection of stimuli given Spanish learners' preference for these two accents as pronunciation models (Mompean, 2004b).

Given the sharp decrease in participation from tweet 22 onwards in Mompean & Fouz-González (in press), the number of target items in this study was reduced from 27 to 22. Most target items were presented in a single tweet, although some were featured in two tweets in order to complete information in the first tweet. This was done in those cases where the researchers suspected that the target word would be unknown for students, offering a translation of the sentence in context, or in those cases where there was not enough space for the link to the audio/video file. In those cases, the second tweet was sent immediately after the first, in order to complete information in that first tweet.

As indicated above, the third study in this dissertation stems from the assumption that the problematic aspects addressed here can be improved through explicit instruction and conscious attention to form, facilitated by input enhancement. There are numerous ways in which input can be enhanced (see Wong, 2005), such as offering metalinguistic information about target features learners fail to notice (Ioup, 1995), or modifying typography in order to make certain features more salient (Han et al., 2008; Saito, 2013). In this regard, study 3 makes use of these two resources in order to draw the participants' attention to the target pronunciations. On the one hand, tweets included concise, explicit, metalinguistic explanations about the problematic features with a link to an audio or video file illustrating the target pronunciation in an authentic context. On the other hand, tweets used, for example, capital letters to signal stressed syllables, or other pronunciation-related familiar words that could be used as examples to illustrate the target aspects (therefore fostering meaningful learning), such as the tweet for *archives* (see Figure 67).

Figure 67. Sample tweet for the target word *archives*

An effort was made to make tweets as simple and student-friendly as possible, drawing the participants' attention to the target features in a straightforward manner. Phonemic transcription was avoided to elude possible problems with symbol displays or adding extra difficulty to the tweets. The links to audio or video files were between one and two minutes long. They contained excerpts from interviews, video clips, news or songs covering a wide range of topics in an attempt to address different learner interests. Videos were downloaded from YouTube or news sites like the BBC or the CNN using Real Player. Files were then edited with Windows Movie Maker so that videos were as short and as little time-consuming as possible, including only the essential information. After the editing process, videos were uploaded onto a private YouTube account so that videos were only accessible to users who had the link.

d) Testing procedure

The participants' improvements in this study were assessed with the oral-matching task explained in section 3.1.3b. There were a total of three production tests in this study: a pre-test, a post-test, and a delayed post-test. The post-test was conducted immediately after instruction took place, the week after training had finished. The delayed post-test was conducted a month after the post-test in order to measure whether gains had been retained over time.

e) Testing stimuli

Testing stimuli consisted of a set of 75 potentially problematic items intermixed with 25 distractors, or items whose stress pattern and grapheme-phoneme correspondences were considered to be uncontroversial for students. Stimuli consisted of three subgroups of 25 words featuring segmental and suprasegmental aspects often problematic for L1 Spanish speakers (see Appendix 26).[108] Groups included silent letters often pronounced by Spanish learners, unusual grapheme-phoneme correspondences, or incorrect placement of lexical stress. More specifically, silent letters included, for example, <b> (*combing, bombing*), <w> (*Greenwich*, *sword*), <l> (*half, talk*), <t> (*fasten, hasten*), or <s> (*aisle, island*). Unusual grapheme-phoneme correspondences included, for example, <ea> in *steak* (often mispronounced as */stiːk/), or <au> in *gauge* (often mispronounced as */gox/, */ˈgauxe/, or */gotʃ/). Items addressing problematic stress patterns included underived lexical items (e.g. *catholic,* often mispronounced as **catholic*), or items whose suffixes play a role in stress placement (e.g. <-ist> in *guitarist*, often mispronounced as **guitarrist*) – for a more detailed account of these aspects, see Mairs (1989) or Monroy (2001).

f) Procedure for the evaluation of production stimuli

Given the relative simplicity of the items under evaluation (e.g. whether students pronounce the <ch> in *archives* as /tʃ/ rather than /k/), stimuli from this study were analysed by only two judges. In the very few cases where there was not agreement between the judges, a third judge was used for disambiguation. Stimuli from the three testing times (pre-test, post-test, and delayed post-test) were edited individually. Each target token was saved as a separate file for its subsequent evaluation. The total number of tokens should be 7,986 (121 participants x 22 words x 3 testing times). However, the total number of stimuli was smaller (6,934) as not every participant completed the delayed post-test.

6.3.3 Results

As mentioned above, the learners' performance in this study was measured with an oral matching task in which learners had to associate words with colours based on their impressions (see section 3.1.3b). Tests consisted of 75 words featuring a number of problematic aspects and 25 distractors. In order to determine the effectiveness of the approach, only the target words were analysed. Since learners received a total of 22 tweets, that was the maximum score possible in the tests.

The aim of this study was not so much to modify the learners' perceptual categories for certain target sounds or to measure their performance when pronouncing certain sound or pattern. Instead, the study aimed to assess whether learners had been able to correct their pronunciation of the problematic aspects addressed. Thus, rather than comparing the

---

[108] This was the same list used in Mompean and Fouz-González (in press).

learners' mean scores at pre- and post-tests, the results only consider that improvements have taken place when a particular stimulus had been mispronounced in the pre-test and was pronounced correctly in the post-test. Learners were given one point per each item they pronounced correctly in the post-test if that word had previously been mispronounced in the pre-test. If they had pronounced that word correctly in the pre-test, no points were awarded even if they still pronounced it correctly in the post-test, as they did not have problems with that word and therefore it did not reflect learning.

In addition, as noted above, a delayed post-test was conducted a month after participants had finished their training in order to check whether they retained the gains obtained from the instruction. In line with the approach adopted for post-test scores, participants were only considered to *retain* learning if they had previously shown *learning*. In other words, they only received points in the delayed post-test if they had mispronounced the word in the pre-test, pronounced it correctly in the post-test, and kept pronouncing it correctly in the delayed post-test. This fine-grained analysis of the learners' progress was considered to offer a more reliable measure of the effectiveness of instruction than simply comparing average scores, as learners may mispronounce a word in pre- and post-tests but pronounce it correctly in the delayed post-test for whatever reason – therefore increasing their *retention* scores even when learning had not occurred.

Since this study did not compare mean scores at pre- and post-tests but the degree of improvement between groups, the data were analysed with T-tests when the data were normally distributed and with Mann-Whitney U tests when it was not. This time, G1 acted as control group and G2 acted as experimental. The interrater reliability for this study was also measured with the Fleiss' Kappa test. The reliability measure was found to be 0.988, an 'almost perfect agreement' (0.81-1.00 range).

Pronunciation improvement

In order to measure the exact effect of instruction, a *percentage of learning* was calculated by comparing the number of words learnt by each group to the total learning potential for each group. Even though the vast majority of participants had problems with the target items addressed, not all participants mispronounced the 22 target items. Thus, the number of words that were pronounced correctly in the pre-test were deducted from the total of 22 yielding a measure of the *learning potential* for each participant. As a case in point, if a participant had already pronounced correctly four words in the pre-test, the learning potential for this participant was 18, rather than 22. This was considered essential in order to provide a reliable measure of the effect of instruction, since two students may have learnt the same amount of words (e.g. 15), but the impact of learning would not be the same for someone who had mispronounced 20 word in the pre-test (75% of learning) than for someone who had mispronounced 15 (100% of learning).

Participants' occasional mispronunciation of other aspects of the target words were not taken into account as long as they had corrected the problematic target aspect (e.g. inserting an epenthetic /e/ at the beginning of *steak* was not penalised as long as the target aspect <ea> was pronounced /eɪ/).

In order to determine the total learning potential for each group, the number of words that were mispronounced in the pre-test was calculated for participants in both groups. That is, the total number of words that could be improved. As explained in section 6.2.2a, there were 43 participants in G1 and 78 in G2. Since there were 22 target words that were considered problematic for the target group, and therefore, likely to be mispronounced, these were multiplied by the number of participants in each group. G1 produced a total of 946 target tokens in the pre-test (22 words x 43 participants) and G2 produced 1716 (22 x 78). Of the 946 tokens by participants in G1, 866 were mispronounced in the pre-test (91.5%). Similarly, of the 1716 tokens by participants in G2, 1577 were mispronounced (91.9%).

Post-test scores show that participants in G2 improved their pronunciation of 413 words (26.2% of their learning potential) and participants in G1 improved their pronunciation of 25 words (2.9% of their learning potential). Moreover, of the 413 words participants in G2 improved from pre- to post-test, 225 were retained a month later (54.5% of the words they learnt and 14.3% of their total learning potential). Of the 25 words participants in G1 improved, two were retained in the delayed post-test (8% of the words they learnt and 0.9% of their total learning potential).

The learning (L) and retention (R) data for groups 1 and 2 were compared with a Mann-Whitney U Test. The results show that the differences between groups reached significance both in terms of learning (U = 366, p = 5.33e-13 < 0.05) and retention (U = 684, p = 0.00000000173 < 0.05). Furthermore, T-tests were conducted in order to assess whether the improvements made by G2 were statistically significant. Both the results for learning ($p = 5.011 \times 10^{-17}$) and retention ($p = 3.447 \times 10^{-15}$) were considered to be significant (Figure 68).

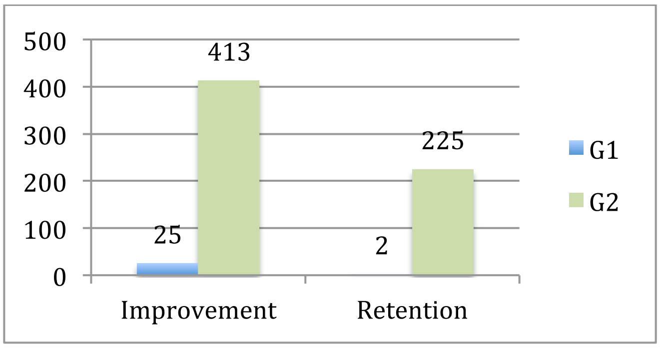


Figure 68. Amount of improvement and retention of learning by participants in G1 and G2

It is worth noting that the progress made by participants in G2 was not homogeneous. There were participants who improved their pronunciation of up to 17 words and participants who did not make any progress at all. As a case in point, focusing on the data for G2, there were three participants who improved in more than 15 words, 16 that improved in nine or more than nine words (including the aforementioned), and 34 improved in six or more than six words (including the aforementioned). The total rate of learning per participant can be found in Appendix 27.

In addition to these measures of learning, the percentage of improvement obtained for each target aspect was also explored. As explained in the training stimuli section, the target aspects included items whose pronunciation was problematic due to sound-spelling correspondences (SSC), silent letters (SL), or incorrect lexical stress (LS). The results obtained for each of these aspects are presented in Table 44.

Table 44. Amount of learning (L), retention (R) and learning potential (LPOT) for each of the target aspects addressed: sound-spelling correspondences (SSC), silent letters (SL) and lexical stress (LS)

	SSC			**SL**			**LS**		
	LPOT	**L**	**R**	**LPOT**	**L**	**R**	**LPOT**	**L**	**R**
G1	324	7 (2.2 %)	0	351	10 (2.8%)	2 (20%)	191	8 (4.2%)	0
G2	569	127 (22.3%)	68 (53.5%)	654	191 (29.2%)	112 (58.6%)	354	95 (26.8%)	45 (47.4%)

Focusing on the results obtained by G2, the data suggest that instruction had a similar impact on the three aspects, all of them showing similar rates of improvement and retention. Participants' improvement rates were 22.3 per cent for sound-spelling correspondences, 29.2 per cent for silent letters and 26.8 per cent for lexical stress. Similarly, they retained 53.5 per cent of the items they improved regarding sound-spelling correspondences, 58.6 per cent of those for silent letters and 47.4 per cent for lexical stress.

One of the findings reported by Mompean and Fouz-González (in press) was that learners' confirmation of tweet-reading decreased considerably after the twenty-second tweet, which led to a reduction of the total number of tweets in the present study (from 27 to 22 tweets). However, despite the fact that this reduction was hypothesised to make instruction less demanding in terms of time, it was interesting to analyse possible differences in the results for tweets that were sent at the beginning of the study and those that were sent at the end. If participation decreases towards the end of the study, the last tweets should show poorer results than those that were sent at the beginning. Nonetheless, the first tweets may also be negatively affected by the fact that more time passes from the moment they are sent to the post-test and may therefore be more easily forgotten than the last tweets. Figure 69 shows the degree of improvement (L) and retention (R) across time, illustrating the number of participants that improved their pronunciation of each target word from the first tweet to the last.

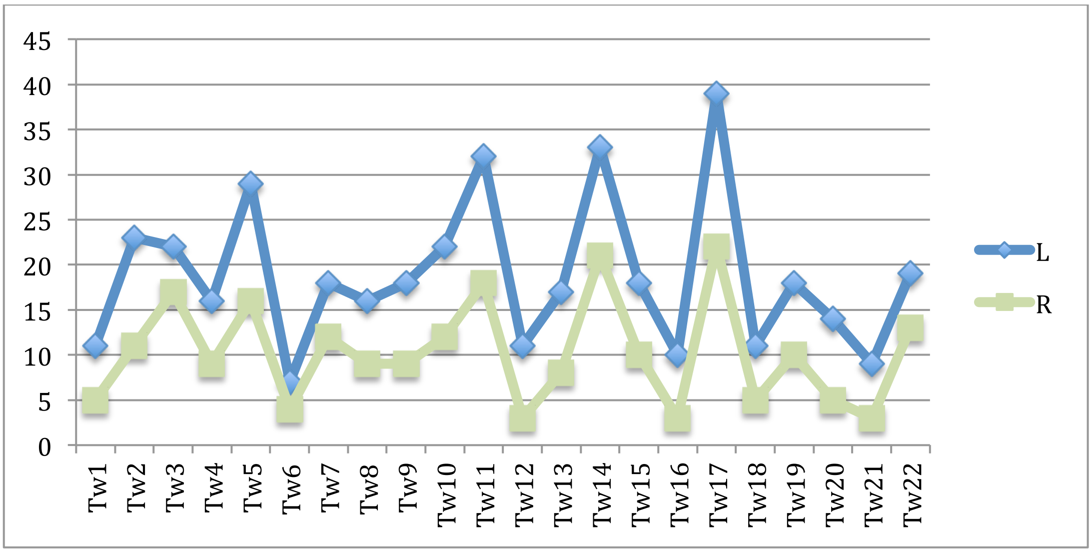


Figure 69. Degree of improvement (L) and retention (R) across time showing the number of participants that improved on each tweet (Tw)

The data suggest that time does not seem to have an impact on either learning or retention, as there are no observable differences among the first and the last tweets. Some of the tweets which obtained the poorest results were sent near the halfway point of instruction (e.g. Tw6, Tw12). Similarly, some of the tweets with the greatest learning rates were sent at the beginning of training (e.g. Tw5, Tw11) and some at the end (e.g. Tw17). As can be seen in Figure 69, the results for retention follow exactly the same pattern.

As explained above, the instructions participants received in this study required them to not only reply to the researcher's tweets confirming that they had read them (as in

Mompean & Fouz-González, in press), but also to write the sentence in which the target word was embedded – which was supposed to provide a more reliable measure of whether learners actually accessed the videos/audios. In our previous study, learners were given one point if they simply replied to the researchers' tweets with 'seen' or 'OK', two points if they provided additional comments (including their own examples, questions, etc.) and no points if they did not reply. This point system was devised in order to measure student engagement in the study, as one of the research questions investigated Twitter's potential to foster participation and participants were not rewarded in any way in that study.

In the present study, since learners were awarded points in exchange for their participation, the scores measuring the participants' engagement were much more demanding. They were given one point only if they replied to the researcher's tweets with the sentence in which the target word was embedded and two points if they provided the word *and* an additional comment, example, etc. However, in this study, learners were given zero points if they simply replied with 'OK' or a similar answer (i.e. not demonstrating having listened to the audio/watched the video), and -1 point if they did not reply at all. The points learners received are presented in Appendix 28 below, together with their *participation score*. These scores were obtained by calculating the percentage of participation for every learner based on the expected participation. That is, receiving one point (if they read the tweet and replied as expected) in each of the 22 tweets.

Furthermore, since some participants in our previous study acknowledged having read several tweets at a time several days after they were sent, a record was made of the time at which learners read the tweets. This was considered as an additional score that measured the learners' punctuality in reading tweets, and therefore, a further measure of participation. In our previous study, participants volunteered to participate purely out of their interest in the language; they were extremely motivated to make the most of any opportunities to learn English. Hence, even if they read tweets late, the researchers considered that they read them because they wanted to and that they were interested in the information they were receiving. Informants in the present study, though volunteers too, were receiving points in exchange for their participation. This may have led some of them to simply reply to tweets in order to get the points at the end of the project. Thus, this further measure of student engagement was considered to contribute to offering a more exact picture the students' participation.

The participation scores for G2 as well as their the *late-reading* scores can be found in Appendix 28. Late reading scores were calculated by adding up all the days that students had read tweets late. That is, if a participant read Tw1 and Tw 2 on the fifth day of instruction, (s)he would have three *late* points from Tw 2 and four *late* points from Tw 1 (i.e. seven *late* points). Since this score reports late participation, it is a negative measure of the learners' engagement. In other words, the higher the students' *late* scores, the worse their participation. As noted above, Appendix 28 also includes the learners' *learning* scores in order to establish comparisons among the different scores.

In order to analyse the data from participants receiving instruction, a convenient classification was to group them depending on the benefit obtained from instruction. Since the maximum *learning* score possible was 22, four broad groups were created: those who improved between 0 and 5 words (poor effect of instruction), between 6 and 11 words (average), between 12 and 16 words (good), and those who improved 17 or more words (very good). This grouping yielded 44 participants in the first group (I1), 26 in the second group (I2), 7 in the third (I3) and one in the fourth (I4).[109]

The data for the only student in I4 (ppt 91) shows a high correlation between the level of engagement and learning, with a participation score (PS) of 118.2%, only 7 negative *late* points and a high *learning* (L) rate (17 words). This is in line with the data from some students in I3. For example, participant 33 has a PS of 95.5%, only 3 late points and a learning rate of 16 words (see also participants 60 or 136). However, there are also students who show considerable improvements despite their low participation scores. As a case in point, three students in I3 (ppts 70, 146 and 158) show quite positive gains from instruction (13, 14 and 14 words respectively) despite their low participation scores (50%, 27.3% and 0%) – see also participants 63, 81, 82, 132. Additionally, there were also a few instances of learners with considerably high *late* scores (e.g. ppts 81, 87, 150) and who nevertheless made substantial improvements from pre- to post-test.

The relationships between learning and participation, learning and late replies, and learning and mere reading confirmation replies were investigated using Spearman's Rho correlation tests. No significant correlations were found between learning and reading confirmation tweets or between learning and late scores (although the *p*-value for the former was very close to the 0.05 significance: $r_s = 0.21$, $p = 0.07 > 0.05$). However, a

[109] In order to avoid possible confusions with the nomenclature for the experimental (G2) and the control group (G1), participants in G2 have been classified into four subgroups with a different label (I) depending on the benefit obtained from instruction (namely, I1, I2, I3 and I4).

positive correlation was found between participation scores and learning. High learning scores were usually associated with high participation scores ($r_s = 0.27$, $p = 0.02 < 0.05$).

Questionnaires

With regard to the learners' responses to the questionnaires sent before and after instruction, 133 participants completed the initial questionnaire and 68 the final questionnaire (24 from G1 and 44 from G2). As explained in section 6.2.2a, the number of participants in this study was 122. There were 12 students who completed the initial questionnaire but then never started instruction. Nevertheless, their responses are considered here given that the data from this questionnaire was meant to gather demographic data from this type of participants as well as their familiarity with Twitter, use of technology, etc.

The learners' responses indicate that they were avid SNS users. Out of the 133 respondents, 115 (86.5%) used SNSs regularly and 18 (13.5%) did not. As for their preferred SNSs, Twitter was used by 76.7 per cent of the students, Facebook by 63.9 per cent, Tuenti (a Spanish SNS) was used by 17.3 per cent, Instagram by 4.5 per cent, LinkedIn by 2.3 percent and Vine by 0.8 per cent.

Regarding the frequency with which participants check SNSs for updates, 29.3 per cent checked them four or more times a day, 27.1 per cent two or three times a day, 15.8 per cent once a day, 6.8 per cent around four or five times a week, 2.3 per cent once a week, and 0.8 per cent once a month. Concerning the amount of time they spend on SNSs, 6.8 per cent spent more than two hours a day, 11.3 per cent spent between one and two hours a day, 18 per cent from 30 minutes to an hour a day, and 51 per cent spent half an hour or less.

One of the open-ended items in the questionnaire asked learners to choose the aspect(s) they considered to be more difficult when learning a foreign language, with grammar, vocabulary and pronunciation as examples. Fifty-five learners mentioned pronunciation (41.4%), 32 mentioned vocabulary (24.1%), 23 mentioned oral comprehension (17.3%), 20 mentioned speaking (15%), three mentioned writing (2.3%), and one mentioned reading (0.8%).

Despite several reminders, the final questionnaire was only completed by 68 participants. Forty-four were from G2, the experimental group, and 24 were from G1. For the items regarding the learners' behaviour during the study, only the responses from participants in G2 are considered. Nonetheless, G1's answers are also included in those addressing general use of the technologies employed or the learners' e-routines.

As regards the frequency with which participants in G2 checked the internet during the study (n = 44), one student claimed that (s)he was online almost all the time (2.3%), 17 claimed to have logged in several times a day (38.6%), 16 once a day (36.4%) and 10 several times a week (22.7%). Six students said that they often logged in in the morning, while the rest mentioned the afternoon or the evening. As for the time they spent on Twitter, two students spent between one and two hours a day (4.5%), seven spent between thirty minutes and an hour (15.9%), six between fifteen and thirty minutes a day (13.6%), and 29 spent less than 15 minutes a day (65.9%).

Twenty-nine of the respondents from the experimental group claimed to be passive Twitter users (65.9%). That is, they did not usually post updates or interact through this SNS but were more keen on simply reading other people's posts. Thirteen students claimed to use Twitter actively (29.5%) and two students mentioned that they did not use Twitter before the study (4.5%). More specifically, four participants said that they sent one or two tweets a day (9.1%), 27 claimed to send less than a tweet a day (61.4%), four said that they sent several tweets a week (9.1%), seven sent several tweets a month (15.9%) and two said that they hardly ever tweeted (4.5%).

As explained above, in order to avoid that learners purposefully studied the target words for the post-test, they were told that they would participate in a study investigating synaesthetic associations between words and colours. One of the items in the questionnaire overtly asked learners whether they had suspected that the study was about pronunciation, and in case they had, they were asked to indicate their participant number in order to omit these data from the analysis. Twenty-seven students said that they had indeed suspected that the study was about pronunciation, 14 said that they did not and two marked the 'other' option and included additional comments. One of the additional comments said that pronunciation was obviously important in the instruction, but that (s)he actually believed it was something the researcher was offering in exchange for their participation in the synaesthesia study. Another comment said that (s)he thought that the goal of instruction was to teach the meaning of the words in the tweets, the way in which they should be spelt, as well as pronunciation. Only one participant claimed to have revised the target words before the post-test (participant 129). Nonetheless, given the low number of points received by this participant in terms of *learning* (i.e. 2 points), the data were not excluded from the analysis, as the fact that (s)he revised does not seem to have affected the post-test scores.

The learners' responses to the post-test questionnaire indicate that they accessed Twitter and read the researcher's tweets from different devices and locations. Forty-nine learners mentioned their mobile phones (18 from G1 and 31 from G2), 40 mentioned their computers (12 from G1 and 28 from G2) and five mentioned tablets (two from G1 and three from G2). As regards the places from which they logged in, 62 mentioned their home (18 from G1 and 44 from G2), 32 mentioned the university (9 from G1 and 23 from G2), two mentioned work (1 from each group) and 21 mentioned some kind of transport (bus, tram, etc. – six from G1 and 15 from G2). Finally, when asked whether learners had felt their privacy to be compromised by sharing their account with their teacher, two students said 'yes' and 65 replied 'no'.

Additionally, several items addressed the learners' perceptions regarding Twitter's potential for education, language learning, as well as several aspects of FL learning. These were presented as five-point Likert-type questions. Since both groups received some kind of EFL-related information during the study, the responses collected from both groups are considered in the analysis (see Table 45).

Table 45. Participants' responses to the Likert-type items in the final questionnarires for study 3 including standard deviations in parentheses

Item	G1 n = 24	G2 n = 44	All N = 68
1. Twitter is potentially useful for educational purposes	3.6 (0.8)	3.4 (1)	3.5 (0.9)
2. Twitter is potentially useful to teach grammar	3.6 (0.8)	3.1 (1)	3.3 (0.9)
3. Twitter is potentially useful to teach false friends	4.2 (0.8)	3.6 (1.1)	3.8 (1)
4. Twitter is potentially useful to teach pronunciation	3.1 (0.9)	3.3 (1.3)	3.2 (1.1)
5. Twitter is potentially useful to teach vocabulary	4.2 (0.6)	3.9 (1.1)	4 (0.9)
6. Because of having received the tips through Twitter, I will be able to remember them better	4 (0.7)	3.5 (1.2)	3.7 (1)
7. The tips I received through Twitter are useful	4.2 (0.7)	4.2 (0.8)	4.2 (0.7)

As can be seen in Table 45, learners seem to perceive Twitter as potentially useful for educational purposes, although their opinions vary depending on the aspect to be taught. The responses from participants in G1 reveal that they consider Twitter to be most useful for teaching vocabulary and false friends, followed by grammar, and finally pronunciation. However, participants in G2 rated vocabulary highest, followed by false friends, pronunciation, and finally grammar.

One of the items asked learners whether they though they would remember the tips they received. Twenty-three participants in G1 replied 'yes' and one replied 'no', while in G2, 37 participants replied 'yes', five replied 'no', and two replied 'sometimes'. When asked about whether they would remember tips better because of having received them through Twitter, the mean scores in the answers to the Likert-type questions was higher for participants in G1 (MS = 4; 0.7) than for participants in G2 (MS = 3.5; 1.2). That is, learners in the control group thought they would remember the tips they received better than learners in G2, the group receiving the pronunciation tips. Additionally, when asked whether they considered reading tips once enough so as to remember the information, four participants in G2 replied 'yes' and 35 said 'no'. Finally, the learners' responses to item 7 in Table 45 above suggest that overall, they regarded the tips they received as useful.

6.4.4 Discussion

The results reported above offer empirical data to provide an informed answer to the research questions for this block.

RQ1

The first research question pertained to Twitter's potential to help learners improve their pronunciation of commonly mispronounced words. It was hypothesised that sending learners a daily tweet with explicit explanations about the target aspects accompanied with audio/video illustrations of pronunciation would help learners notice the problematic features in the target items and would consequently lead to improvements in their pronunciation of those words. As noted above, participants in the experimental group (G2) improved their pronunciation of 413 words (26.2% of the words they mispronounced in the pre-test), which was considered significantly different from the minimal improvement made by G1, which was of only 25 words (2.9% of the total they mispronounced in the pre-test). Furthermore, the results obtained in the delayed post-test reveal that instruction not only helped learners correct their pronunciation of the target aspects, but participants also retained this 'learning' a month after instruction. Participants in G2 retained 225 words (54.5% of the words they learnt and 14.3% of their total learning potential) while participants in G1 retained only two (8% of the words they learnt and 0.9% of their total learning potential).

Research has shown that Twitter can help learners improve their pronunciation of a number of problematic aspects (Fouz-González & Mompean, 2012; Mompean & Fouz-González, in press). Nevertheless, one of the limitations of these two studies was the lack of a control group offering a higher reliability to support that improvements in the pronunciation of the target words were due to the type of instruction and not because learners had encountered these words elsewhere. The substantial differences between the experimental and control groups in the present study suggest that improvements from pre- to post-test were due to the Twitter-based instruction. The improvement by learners in the control group was very modest. Moreover, the fact that of the 25-word improvement by participants in G1 only two were retained after a month suggests that their adequate pronunciation of those words in the post-test may have been incidental.

Additionally, the above-mentioned studies used a rather limited number of participants, with only eight students in the first (Fouz-González & Mompean, 2012) and 16 in the second (Mompean & Fouz-González, in press). In this regard, the present study offers a considerably larger sample, with 42 students in the control group and 79 in the experimental group (i.e. 121).

The high percentage of target words that were mispronounced in the pre-test (91.5% by participants in G1 and 91.9% by participants in G2) confirms the researcher's predictions regarding the problems these items pose for EFL learners. Even though the final items used for the present study were not exactly the same as the ones used in Fouz-González and Mompean (2012) or Mompean and Fouz-González (in press) – as they were selected after an analysis of a sample of 15 pre-test productions – they are largely the same target words that were used in these prior studies.

The fact that the percentages of improvement and retention were similar for the three types of aspects addressed (sound-spelling correspondences, silent letters and lexical stress) suggests that improvements were made regardless of the target aspect. The aspects addressed here and in previous studies are considered to be problematic due to a lack of attention to form which could be overcome by explicit instruction and input enhancement that should facilitate noticing. The results reported above provide empirical data suggesting that these aspects can be modified with form-focused activities and explicit instruction.

Additionally, the results obtained for each tweet were compared in order to explore possible differences in the effectiveness of instruction depending on the time of the study at which the tweet was sent. It could be the case that the last tweets received lower scores than the first if participants stopped reading tweets as a result of the duration of the study. Another possibility was that learners forgot the first tweets they received and remembered better the ones that were sent towards the end of instruction. However, the data suggest that time did not affect learning or retention, since there were no observable differences among the first and the last tweets.

With regard to the participants' engagement in the study, measures of participation were much stricter than the ones used in previous studies (see above). Participants in Mompean and Fouz-González's study volunteered to participate in their study and were not offered anything in exchange for their participation. Thus, the researchers had no reason to question whether the learners' reading-confirmation tweets were reliable confirmations that they had actually read the tweet. Nevertheless, since participants in this study were awarded 0.3 points of the final mark if they participated in the study, there was a risk that they claimed to have read the tweet even if they had not done so. For the same reason, they were also given a *late* score in order to control whether they read the tweets as they were sent or whether they read several tweets together. A detailed inspection of the different participation measures revealed several contradictions. For example, there were learners with very low participation scores who nonetheless showed considerable improvements from pre- to post-test. This suggests that learners may have participated actively and read the tweets despite not having followed the researcher's instructions regarding the way in which they had to confirm reading them. That is, learners may have simply replied to tweets without including the sentence in which the target word was embedded. Likewise, there were several participants who made substantial improvements from pre- to post-test with considerably high *late* scores. Despite the existence of some contradictory patterns, the results from a Spearman's Rho Correlation test revealed that high learning scores were usually linked to high participation scores. Therefore, even if the scoring system devised does not serve to explain certain participants' behaviour, it seems that it was an accurate reflection of the learners' engagement in the study.

The fact that there were several participants with high *late* scores but also high *learning* rates suggests that even if students are willing to participate in this type of instruction, some may prefer to access materials at a time of their convenience. Given the bite-size information shared in this study, perhaps some students prefer reading several tweets at a time. In the study by Saran et al. (2009), learners received four items a day via MMS and most students considered it appropriate. However, this would defeat the original purpose of this approach, which consists in directing the students' attention to just one item at a time. Even though students may occasionally read several tweets at a time due to their busy schedules, for the approach to be a way of enhancing input, sending a 'word of the day' is considered enough. This word of the day is salient precisely because it is 'the word of the day'. Sending more than one item per day would reduce the saliency of the target aspects.

RQ2

The second research question was whether Twitter can also be effective with learners who volunteer to participate than with students who are rewarded for their participation. It was hypothesised that if this type of Twitter-based instruction training yielded positive results with a group of volunteer FL learners, training with a group of students enrolled on an ESP course and rewarded for their participation should yield similar results.

The findings in Mompean and Fouz-González (in press) revealed that when this type of Twitter-based pronunciation instruction was implemented with highly motivated students, the mean gain rate was of 75.2 per cent. Even though not every participant obtained such a high improvement, with four students showing an improvement of less than 65 per cent, the benefits obtained from this prior study were much greater than the ones obtained in the present study, where the mean gain rate was 26.2 per cent. The fact that this study did not have such a positive impact on the learners' pronunciation of the target words does not mean that instruction was not effective. As noted above, participants in the experimental group in this study made significant improvements from pre- to post-test, retained them a month after training, and these improvements were significantly superior to the ones made by the control group.

A close inspection of the data reported above and the one obtained in Mompean and Fouz-González (in press) points to the level of student engagement as one of the main reasons explaining the lower gain rate obtained in the present study as compared to the previous one. As commented above, the participation scores in this study were closely linked to the benefits obtained from instruction. In this regard, the participation rate in the present study was only 30 per cent (close to the 26.2% of benefit) as compared to the participation rate obtained in our previous study, which reached 81.8 per cent (close to the 75.2% of benefit for that study). This suggests that the higher the learners' active engagement in the instruction, the higher the benefits they obtain from it.

Focusing on the idea underlying the approach, that is, whether Twitter is effective in helping learners modify aspects that they commonly mispronounce, the data suggest that it is. The results from the present study show that participants were able to improve up to 17 words out of 22 (77.3%) thanks to the Twitter-based instruction. Put simply, Twitter *can* help learners improve their pronunciation of words that are commonly mispronounced due to lack of attention to form. In fact, if the results from the 10 students who obtained the best results from this study are isolated and compared with the results in our previous study – as production data were obtained from 10 students in that study –, the mean of improvement is 69.1 per cent, which is similar to the improvements made by participants in the previous study (75.2%).

The above discussion suggests that Twitter can be effective in order to help learners improve their pronunciation of aspects that are problematic for FL learners, although the benefit that can be obtained from instruction is subject to their level of engagement. The different levels of participation in these two studies may be due to a higher motivation to learn English by students in Mompean and Fouz-González's (in press) study. These learners were studying English as something extra-curricular and completely voluntary. They were enrolled in a six-year period of EFL instruction at an official language school. However, participants in the present study were enrolled in a compulsory ESP course as part of their degree in Medicine. While the former group's willingness to participate in the study is a clear indicator of their eagerness to learn, the latter group may have been driven by the external reward offered.

In line with the other studies in this dissertation, participants were not awarded points depending on how well they completed each individual task, but on task completion. In the case of this study, they were awarded the 0.3 points if they attended the three tests (pre-, post- and delayed). The lower participation rates by learners in study 3 (see Appendix 28) suggest that most learners took part in the study simply in order to get the reward. Even though this could be considered a limitation, it can also be interpreted as a genuine measure of Twitter's potential to foster participation and pronunciation improvements in a group of average FL learners who have no particular motivation to learn the language. Nevertheless, the nature of the present study limits the validity of this statement, as learners were aware that they were taking part on an experimental study and their performance may have been negatively affected as a result of it. Had the Twitter-based instruction been implemented as part of their ESP course and learners knew that items featured in the tweets could appear in exam questions, the results could have been different.

RQ3

The last research question was concerned with the students' reactions towards the Twitter-based instruction (initial and final questionnaires can be found in appendices 29 and 30).

Participants considered the tips received to be useful, although the perceived potential of Twitter for educational purposes varied depending on the aspect to be taught. Pronunciation received the lowest score by participants in G1. They rated false friends and grammar as the aspects for which they perceived Twitter to be most useful. This is not surprising given the tips they received mostly featured these two aspects, whereas pronunciation is not a competence one expects to be improved through messages of 140 characters at first. However, even though grammar was the competence that received the lowest scores by G2, the rating they gave to pronunciation was not much higher than the one by participants in G1.

Most participants believed that they would remember the tips they received through Twitter, which is in line with the positive retention rate revealed by the delayed post-test scores obtained by participants in G2. Nevertheless, the mean score in the responses by G1 was slightly higher than that by participants in G2. This suggests that participants considered vocabulary, grammar and other course-related tips to be easier to remember than pronunciation tips. Additionally, the majority considered that reading tweets only once was not enough. Perhaps sending tweets several times during the instruction could foster bigger improvement and higher retenion rates.

Twitter's potential to improve pronunciation finds support in the data reported in this and previous studies, although as noted above, these studies have focused mostly on problems arising from a lack of conscious attention to form. It is difficult to imagine how Twitter per se (without links to other activities, websites, audios or videos) could help learners acquire other aspects of the FL phonology, such as new phonetic categories, adequate rhythmic patterns, etc.

The learners' responses to the questionnaires reveal that most of them used SNSs regularly and that the majority were already familiar with Twitter. Furthermore, they spend a considerable amount of time online every day. Thus, reading the researcher's tweets amidst their e-routine should not pose excessive time demands. However, while 18 students said they logged in several times a day, 16 only logged in once a day and 10 only several times a week. The idea underlying this approach is that learners can read the researcher's tweet while checking the updates by other accounts they follow, although for the group logging in only several times a week, accessing Twitter daily may have not been so convenient.

The above suggests that Twitter can be useful for pronunciation instruction, at least for aspects that do not depend on the learners' perceptual or articulatory abilities, but on their lack of conscious attention to form, lack of familiarity with the phonological composition of words, or with certain sound-spelling associations. Nonetheless, there are several limitations to the present study that must be acknowledged and which offer directions for future research.

Limitations and directions for future research

The first limitation is the fact that tweets remained in the researcher's account during the whole instructional period. That is, if someone could not read the tweet on the day it was sent, (s)he could go to the researcher's account and check tweets from previous days. This was done in order to make it possible for participants to access tweets even if they had to read them late. This was not considered problematic insofar as participants believed the synaesthesia pretense. As explained above, participants were told that they would help the researcher carry out his investigations with sound-colour associations and that they would receive tips about English in exchange. The researcher informed participants about the existence of previous research investigating sound-colour associations focusing on English vowels (see e.g. Mompeán-Guillamón, 2014) and its application to language teaching (Mompeán-Guillamón, 2015). Participants were told that the goal of the study was to investigate whether the colour with which they receive each tweet (in the researcher's profile picture) could affect they synaesthetic perception of the words as compared to their initial perceptions in the pre-test (see Figure 67). This pretense was maintained until the end of the study by changing the colour of the researcher's profile picture every two days. If learners believed this to be the researcher's goal, there would not be a reason to try to memorise the pronunciation of the target words for the post-test. Their responses to the final questionnaire reveal that they indeed suspected that pronunciation was an important part of the study, but not more important than vocabulary, idioms, or peculiar spellings, as revealed by the comments to the open-ended questions. Furthermore, they denied having studied for the post-test, except for one student who made an improvement of only two words.

Despite the fact that care was taken in order to make the study seem to deal with something else, perhaps it would be convenient to delete tweets the day after they were sent, although this would not be free from limitations either. On the one hand, learners would not be able to access tweets late, something most students did in the present study given that Twitter allows users to access content at a time of their convenience. On the other hand, this would not ensure that learners do not study for the post-test either, given that if they suspect that they need to learn the pronunciation of those words for the post-test, they could always write the words down when they read them and then look them up in a dictionary.

Another limitation is the specific nature of this study, as learners in G2 may have perceived the words to bear little relation with the course contents. Future studies should explore whether learners obtain better results from this type of instruction when it is integrated as part of their course (e.g. dealing with course-related vocabulary, idioms, or content), even if they are not evaluated based on task completion. Research has shown that when Twitter is used in order to complement course contents, the majority of students are willing to follow the teacher's tweets, even if learners do not post actively (Lowe & Laffey, 2011). Moreover, when used as a way to send reminders about important class concepts, Twitter has been found to help learners remember these better (Blessing et al., 2012). Thus, future research should explore whether this type of Twitter-based pronunciation instruction is more effective when focusing on course contents. In the case of the ESP course for Medicine students, the results might have been better if instruction had focused on the pronunciation of Medicine-related items, for example.

Related to the above is the fact that participants may have simply participated in the study because of the reward. As in studies 2 and 3, some learners' pre- and post-test recordings reveal a clear lack of interest in what they were doing. The results obtained proved to be significant in spite of the low level of engagement by some participants. Nonetheless, this limits the validity of the evaluation of these tools' potential, as it difficult to make something work when learners do not even make an effort. Thus, it would be interesting to conduct more studies in which training is an inherent part of the course, perhaps without letting participants know that their performance is being controlled for.

Previous studies have called for empirical engagement measures in Twitter (Junco, Heibergert, & Loken, 2010), as some research has been mostly concerned with self-reported measures of participation. This was done in Mompean and Fouz-González (in press) as well as in the present study, by measuring the learners' responses to the researcher's tweets and assigning them points depending on the amount of effort required. This was very convenient in our previous study with 16 participants. However, as noted by Rinaldo et al. (2012), checking participation when students are required to post can be extremely time consuming for large groups. Research has shown student engagemet and grades are better when faculty interacts with them regularly (Junco, Elavsky, & Heiberger, 2013). Therefore, the researcher spent several hours a day keeping a participation log as well as replying to the 121 participants in this study. Each participant received an individual response via private messages every day as a response to their reading confirmation tweets (irrespective of the type of confirmation). Although a model answer was created for most tweets so that it could be copied and pasted for every participant (e.g. 'Well done! Thanks!'), the researcher often had to customise individual replies when learners asked questions or commented on the tweets. An online application that checked tweet-reading automatically would make the researcher/teacher's task incredibly easier.

Furthermore, even though the above-mentioned measures are based on learners' actual participation in the study, and despite the effort made to control that they really watched the content (by asking them to write the sentence where the target word was embedded), this is not a perfectly reliable measure that they actually watched the video. In the same way that some students cheat in exams, participants in a study of this kind might well send the carrier sentence to each other via private messages and simply copy and paste each other's answers without even reading the tweet. This could affect the results by students in the experimental group who might obtain poor scores in the post-test not because the Twitter-based instruction is not effective, but because they do not read the tweets – perhaps simply pretending to participate in the study in order to obtain the 0.3 points. This is a limitation that cannot be overcome when using this type of tool. Other tools, such as virtual learning environments at universities allow teachers to check whether learners download course contents (which does not imply that they read it either). However, tools that were not originally conceived for educational purposes do not offer the opportunity to check whether participants read the information they receive. As a case in point, YouTube allows the account owner to see how many times a video has been watched; nonetheless, it does not provide information regarding who. Future research could explore alternative ways of measuring tweet-reading.

It would be worthwhile to investigate whether reading tweets only once is enough so as to assimilate the information in the researcher's explanation, or whether more times are needed – and if so, possibly measuring the number of times participants need to be exposed to the target features. In their responses to the questionnaires, participants claimed that reading tweets once was not enough in order to remember the tips. Thus, it is possible that they checked them several times during the day, or that they played the videos/audios several times. It would have been interesting to measure the exact number of times that participants watched the videos or listened to the audios when they read the tweets.

Nevertheless, this was not controlled for in this study given that neither Twitter nor the video/audio hosting services used offer this feature.

Additionally, despite the fact that learners retained the improvements made in the around half of the words they learnt, this rate could have been higher. It is important to note that not all participants completed the delayed post-test given that they did not live in Murcia and it was impossible for them to attend the last interview just before the exam period. Thus, these data cannot be considered in the scores for retention, offering only a partial sample of the learners level of retention in the delayed post-test.

It must be noted that although questionnaires asked participants about the location and device where they read the tweets, this information was not controlled for each individual participant in this study. Besides, some students responded that they read some tweets at home, some at university, and some while travelling to and from home. Future research could explore whether the place or device used to access content has an impact on the degree of learning and retention.

Finally, given the learners' positive perceptions regarding Twitter's usefulness to learn vocabulary, idioms and other language-related content, future studies should also address the possibilities offered by this tool in order to teach those aspects.

In line with Fraser's (2001) assumption about conceptualisation of English phonology being one of the most challenging aspects for learners, the results obtained in studies 2 and 3 suggest that one of the most common problems learners face is not only to learn how to pronounce a sound, but learning when to use the sound. In this regard, Twitter seems a promising tool in order to raise learners' awareness of the different spellings that may feature English sounds as well as to help learners become familiarised with the phonological composition of words that are problematic.

Chapter VII: Conclusion

7.1 Training FL pronunciation with technology

As discussed in chapter I, technology can enhance pronunciation training in ways that traditional instruction cannot. Therefore, in the same way that language teachers are considered to be indispensable and should never be replaced by a computer, technology should also occupy a central position in language teaching and learning given the wide range of possibilities it offers (Thomson, 2011). Hovever, one of the most cited claims in the CALL/CAPT literature is the mismatch between technology and pedagogy, with software often incorporating flashy features to impress users despite their limited effectiveness for language acquisition (Levy, 1997; Neri et al., 2002; Pennington, 1999). Hence, before making recommendations to implement new tools in FL instruction, research should assess their effectiveness for the acquisition of FL pronunciation empirically.

As noted in chapter I, for technology to really enhance training, activities should genuinely exploit the capabilities of the tools being used, not simply 'rehash' old techniques and do with technology the same we could do without it (Colpaert, 2004; Setter & Jenkins, 2005). Thus, research should investigate the extent to which particular tools can be suitable for a given group of learners at a given point in time (Jamieson et al., 2005). In this regard, in the three studies in this dissertation, an effort was made to make the most of each technology, exploiting the capabilities offered for the aspects they were considered to be most useful.

The two studies described in chapter IV used podcasts to draw the participants' attention to pronunciation by means of three weekly tasks. First, learners were required to pay careful attention to the target features in the sample podcasts by means of focused listening. Second, they recorded their own podcasts with numerous instantiations of the target features, therefore offering opportunities for productive practice. Finally, they were asked to evaluate their peers' podcasts in order to raise their awareness of pronunciation and foster the development of self-monitoring skills.

Chapter V focused on the potential of smartphone apps for autonomous pronunciation training. The EFP app was selected for perceptual training given that it uses phonetic symbols in order to familiarise learners with the phonological system of English. It is believed that providing learners with these labels should help them classify future instantiations of the target sounds and develop adequate pronunciation targets. Eventually, this should, in turn, help them monitor their pronunciation by comparing their production to the model they have in mind.

Finally, chapter VI addressed the use of SNSs for pronunciation training. Given Twitter's conciseness and the possibility to send free messages with links to audio/video, this SNS was considered to be potentially advantageous for sharing brief explicit tips about key aspects of pronunciation whose difficulty is believed to reside in lack of attention to form. The fact that each tweet was devoted to only one pronunciation aspect was considered to be a way of making input more salient and facilitate learners' noticing.

As mentioned above, the studies in this dissertation represent an attempt at finding empirical evidence on the usefulness of different affordable and easily accessible technologies for FL pronunciation training. However, since researchers have long emphasised the need to move beyond laboratory-like, controlled environments in CAPT (Lord, 2010; Olson, 2014; Wang & Munro, 2004), an effort was made to make studies as unrestricted as possible. This emulates the way learners could use these tools were the teacher to implement them as a complement to classroom instruction. Hence, except for the pre- and post-tests, everything in studies 2 and 3 was done off-campus. Nonetheless, given the findings obtained in the pilot study with podcasts, and due to the impossibility of controlling the completion of some of the tasks in study 1, some of the activities in that study were completed face-to-face.

In order to test the potential of the above-mentioned tools, a number of aspects that are considered to be challenging for FL learners were selected. Study 1 addressed podcasts' potential to improve the learners' perception and production of the English /s – z/ contrast and of /b d g/ as stops in intervocalic position. Study 2 concentrated on the potential of the EFP app to help learners improve their perception and production of the English vowels /æ/, /ɑː/, /ʌ/ and /ə/, and of the /s – z/ contrast. Finally, study 3 weighed Twitter's potential to improve the learners' pronunciation of lexical items that are often mispronounced by EFL learners due to difficult sound-spelling correspondences or cognates in English and Spanish with different stress patterns. The target aspects addressed in studies 1 and 2 are often fossilised in the interlanguage of advanced EFL Spanish learners (Monroy, 2001), and therefore, they were considered to be difficult to modify without explicit instruction. The targets in study 3 are not considered to stem from perceptual or articulatory difficulties, but from lack of exposure to the target items, lack of conscious attention to form or explicit instruction, or a combination of these.

7.2 Findings

The results obtained in the three studies in this dissertation offer empirical evidence of the convenience of using these tools for FL pronunciation training. The data in these studies show that instruction had a positive impact on the learners' perception (studies 1 and 2) and production (studies 1, 2 and 3) of the target aspects. Nonetheless, although the improvements made by the experimental groups reached statistical significance for some of the target features, the results were not so positive for every aspect.

Previous studies exploring podcasts' potential for FL pronunciation training have found beneficial effects of using podcast-based instruction on the learners' general pronunciation ability and had called for research addressing the impact of this kind of instruction on specific pronunciation aspects (Lord, 2008). Nevertheless, other studies investigating the potential of podcasts to help learners improve comprehensibility and accentedness were not very encouraging, even after a period of 16 weeks (Ducate & Lomicka, 2009). As the authors point out, this may be due to a lack of focus on form, as learners in their study tended to concentrate on meaning rather than on pronunciation. In this regard, study 1 offers further empirical evidence of the convenience of using podcasts for pronunciation instruction.

The results obtained in study 1 suggest that a podcast-based instruction with form-focused activities and explicit instruction can foster improvements in the learners' perception of the target aspects, even after only three hours of instruction. However, it is important to point out that the results only reached statistical significance for the /s – z/ contrast, but not for /b d g/. Regarding production, instruction had a significantly positive impact on the learners' pronunciation of /z/ and /g/, and participants generalised their improvements on their production of /z/ to novel words. Additionally, the participants' responses to the questionnaires indicate that they found the approach useful. Their answers regarding their self-perceived progress show that they considered the podcast-based instruction to have helped them improve their perception and production of the target aspects, which is in line with some of the results obtained. Even though the improvements made did not reach significance for all sounds in all tasks, participants in the experimental group generally outperformed participants in the control group.

With regard to mobile apps, previous studies have found that training with a mobile version of an HVPT paradigm can help Japanese learners of English improve their perception of the /r – l/ contrast (Uther et al., 2007), and that speech-to-text ASR feedback can help learners improve their production of French /y/ (Liakin et al., 2014). The results obtained in study 2 offer additional support of the effectiveness of smartphone apps for pronunciation training.

The data from study 2 suggest that training with the EFP app helped learners improve their perception of the target sounds substantially for all groups when they received training. Nonetheless, the analyses of the scores for each sound revealed that even though the three groups made considerable progress after they received instruction, these only reached statistical significance for some groups. Significant developments were found in G1's perception of /ʌ/, /ɑː/, /ə/ and /z/, and G3's perception of /ə/. However, these findings must be interpreted with caution given the possible effects of lexical familiarity in the learners' scores for familiar items. Regarding production, the app-based instruction yielded significant improvements in G1's total scores for /æ/, /ʌ/, and /ə/, in the imitation of /æ/, and in the spontaneous production of /ə/. Moreover, participants in G1 generalised their improvements for /ə/ to novel words in production. It is important to point out that despite the fact that improvements mostly reached significance for G1, all groups showed substantial progress in almost every task when acting as experimental.

Finally, study 3 extends previous research investigating Twitter's potential for pronunciation training. The results obtained in study 3 are considered to be very positive, as participants receiving instruction through Twitter improved their pronunciation of the target words significantly and they retained this improvement a month after training. The benefits obtained from instruction are not as encouraging as those by previous studies (Mompean & Fouz-González, in press), but the level of engagement in this study was also comparatively very low. Learners who participated actively in study 3 did make similar progress to the ones in that previous study, which suggests that provided that learners are willing to learn, this type of Twitter-based instruction can foster considerable developments. Even though participants perceived the tips they received as useful, and despite the fact that they made significant pronunciation gains, they perceived Twitter as potentially more useful for teaching vocabulary or false friends than for teaching pronunciation.

Overall, the findings obtained in the three studies in this dissertation are interpreted to be encouraging. Even though instruction through these tools was not equally beneficial for every target aspect, substantial improvements were made in every study despite the extremely short period of instruction. This is considered to be positive taking into account that the aspects addressed in studies 1 and 2 tend to be fossilised in the interlanguage of advanced FL learners and are, therefore, considered to be very difficult to improve.

Notwithstanding the above, the overall learners' capacity to perceive and produce the target sounds was not modified substantially as a result of instruction, as evinced by the improvements made in the discrimination and imitation tasks, which were similar for the control and experimental groups. With the exception of /b d g/ (which were already very high at the beginning of instruction), the learners' scores were often far from the maximum ratings in the different tasks. In other words, there was much room for improvement in the learners' pronunciation of the target sounds. The same was observed in the results for controlled (sentence-reading task) and spontaneous production (timed picture-description task). Even though the overall production scores reached significance for some sounds, the mean scores for all groups were far from the maximum scores for each task.

A noteworthy finding from the data obtained in studies 1 and 2 is that, with the exception of /b d g/, the learners' pre-test scores in perception tasks were usually higher than those in the imitation task. As noted above, imitation tasks are often used in order to measure phonetic category formation. However, this task taps on both domains: perception and production, given that in order to articulate the sound, participants have to perceive it first. Thus, the data in studies 1 and 2 offer support to the widely acknowledged claim that perception precedes production, as participants were capable of discriminating and identifying the target sounds more accurately than attaining them in production.

7.3 Limitations

This dissertation presents several shortcomings that must be acknowledged. One of the most important limitations of the three studies is the fact that the duration of training was relatively short. Studies 1 and 2 were conducted over a period of three weeks. Study 3 over five. Nevertheless, the amount of instruction participants in studies 1 and 2 received added up to a total of around three hours. This is considered to be enough to test the potential of the resources under study for pronunciation instruction, as it is possible to appreciate whether improvements take place in the perception and production of the target sounds studied. Nonetheless, for learners to be able to pronounce adequately in controlled and spontaneous production and transfer gains to novel words, longer periods of time are probably necessary.

Additionally, learners were given academic credit in exchange for their participation and they knew that they were taking part in research studies. This is considered to have affected the possibilities of the approaches adopted here negatively, as participants knew that what they were doing was something extra and that they would not be evaluated based on how well they completed the tasks, but only on task completion. Hence, some participants may have completed tasks in these studies simply because they would be rewarded, without trying to make the most of the training. In fact, this was appreciated in some learners' lack of interest in pre- and post-test recordings, or in the fact that some of them did not complete the initial and final questionnaires. This limits the validity of the evaluation of the tools' potential for pronunciation training, as some participants were not really trying to learn. The studies in this dissertation were an attempt to move away from the common laboratory-like, controlled situations in other CALL studies. However, the limitation described above suggests that more studies should be conducted to gain further evidence on the potential of the tools under study, perhaps alternatively implementing instruction as part of the learners' course requirements.

A further limitation of the studies in this dissertation is that none of them encouraged oral interactions among participants and the focus in all the activities was on form, not on meaning. The reason for this is that when learners have to process form and meaning simultaneously, the amount of attention that can be devoted to form is minimised, as meaning is more important for communication (Thomson, 2011). Therefore, when the learners' attention is directed to phonetic form, it is easier for them to assimilate phonetic information (Guion & Pederson, 2007). The studies in this dissertation placed a great emphasis on form. Nevertheless, future applications of these tools for pronunciation training could try to direct the learners' attention to form while encouraging more communicative activities and interactions between participants (see Isaacs, 2009; Trofimovich & Gatbonton, 2006).

Finally, the fact that there was so little variation in the training stimuli in study 2 makes it difficult to talk about the potential of smartphone apps for pronunciation training, especially since the study evaluates the potential of only one app, and the app presents several limitations.

7.4 Directions for future research

The findings obtained in the three studies reported above also offer directions for future research. For example, although lexical stress was addressed in study 3, the studies in this dissertation focus almost exclusively on the training of segmental features. As explained above, the selection of the target aspects for each study was based on the fact that they are considered to be very difficult to modify as they tend to be fossilised in the interlanguage of advanced EFL learners. This was very convenient in order to test the effectiveness of the three tools explored empirically, as evidence that aspects as difficult as these can be modified with these affordable and easily accessible tools would provide an alternative to some of the technologies reviewed in section 1.3. Nonetheless, future research should investigate the potential of these tools to train suprasegmental features of speech, especially due to the role they play in intelligibility (see chapter I).

Furthemore, given the numerous claims in favour of mobile learning, it would be interesting to explore the potential of these (podcasts, apps and twitter) and other tools for learning 'on the move'. Learners in the three studies in this dissertation used these technologies from various locations, but these were not controlled for. Future studies could investigate whether perceptual training can have the same impact on the learners' pronunciation when conducted in a quiet environment, such as the learners' homes, than when practising on the street, given the different demands of attention or background noises learners may find when using the app on the go.

Additionally, as pointed out above, the studies in this dissertation focus on a particular smartphone app and on one SNS. Nevertheless, in order to make claims about the potential of these tools for pronunciation training, more studies need to be conducted. Research should investigate the potential of other apps and other SNSs. The EFP app was chosen for study 2 given its high compatibility with all kinds of devices and because the approach adopted was considered helpful in order to help learners become familiarised with the sound system of English. However, it would be interesting to explore the potential of other commercial apps developed by well-known professionals, such as *Sounds* (Macmillan Publishers Ltd., 2011) by Adrian Underhill, or *Clear Speech* (CUP, 2011) by Judy Gilbert. Likewise, study 3 focused on Twitter given its suitability for directing the learners' attention to the target aspects addressed, but it would be interesting to explore the potential of other SNSs for pronunciation training.

Despite the enormous potential of some of the existing apps for pronunciation training, in order to control the participants' development more reliably, researchers should be able to design their own materials and test the potential of a self-developed app grounded on the findings obtained in pronunciation acquisition research. This would not only make it possible to overcome some of the obstacles encountered in study 2, but it will also enable researchers to control the exact amount of practice necessary in order to master key sounds or patterns.

It is important to note that it is extremely difficult for someone without programming skills to develop their own materials without the help of specialists. In fact, this was the original intention for study 2 in this dissertation. One of the researcher's initial goals was to create an app in order to test its effectiveness empirically. At the initial stages of this dissertation, a survey of existing pronunciation apps was conducted with the intention of creating a product that overcame some of the limitations in current apps and testing its potential empirically as part of this dissertation (a brief 'state-of-the-app' was presented at the 2012 EUROCALL conference – Fouz-González, 2012).

In light of the very poor quality of some existing apps (Fouz-González, 2013) and the often cited divorce between technology and pedagogy (Neri et al., 2002; Pennington, 1999), a 'call for collaborators' was sent to all students of computer science (CS) at the University of Murcia. Given the absence of funding, the idea was to find someone who needed to design an application for their Master's degree or PhD programme; the researcher would work on the pedagogical component of the app and the CS specialist would work on the technological one. No candidates were available for such collaboration, but a CS student from the University of Murcia showed his interest in the project and is currently working with the researcher and his two PhD supervisors in an app that will hopefully be released in the near future.[110]

The app the researcher is currently working on would respond to some of the needs stated by participants in study 2, such as different levels of increasing difficulty (Figure 70),

[110] The people involved in this project are Francisco Campillo Asensio (CS) and my two PhD supervisors (Rafael Monroy Casas and José Antonio Mompeán González). It should be noted that the app is still in its early stages, but updates will be posted in the researcher's website: https://sites.google.com/site/jonasfouz/

a much wider set of stimuli featuring sounds in different positions, by different speakers and in a wider range of lexical items (Figure 71, left), or the possibility to listen to words in sentences and short excerpts (Figure 71, right).[111]

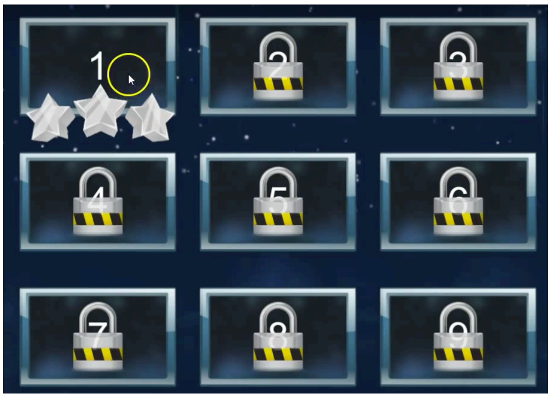


Figure 70. Sample screenshot of a progress screen in the researcher's app

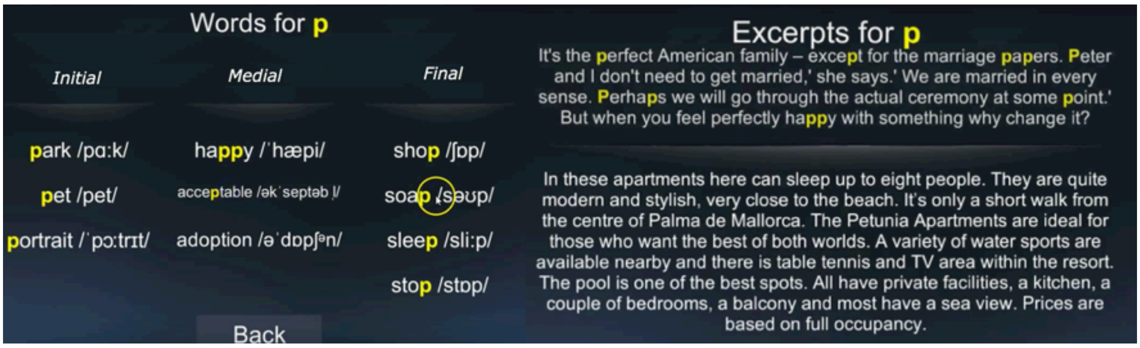


Figure 71. Sample screenshot of /p/ in different positions (left) and in short excerpts (right)

The researcher's app is not intended to concentrate on one single pronunciation aspect. It will include practice on different segmental and suprasegmental features. The app will also have a phonemic chart in order to help learners get to know the phonemic symbols (Figure 72, left), but the idea is to also incorporate relevant information as to why the sound is problematic, possible spellings for the sound, cross-language comparisions, and animated illustrations of sound articulation (Figure 72, right). The app will include several activities to work on segmental identification, but also others aimed at practising word and sentence stress, contrastive stress, or the attitudinal meanings of intonation.

[111] The graphics illustrating the researcher's app are only a preliminary version of what the activities will look like. The programmer is using standard templates in order to test the functions in the app.

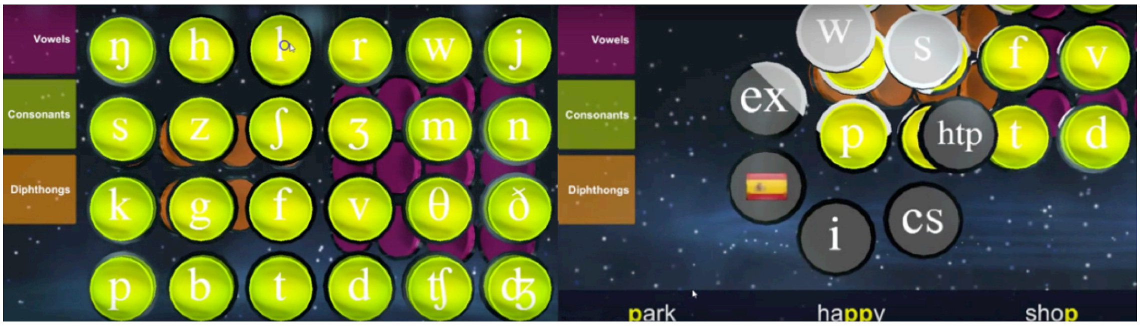


Figure 72. Sample of the phonemic chart (left) and the menu for each sound (right) in the researcher's app

The activities in the app the researcher is working on makes use of various gamification techniques. For example, in one of the activities addressing word and sentence stress (*Train stress*), words are divided into syllables that are represented by wagons from a train. Users are required to place a group of heavy rocks in the wagon carrying stress. If they do it right, the train passes through a tunnel; if they fail, the train is derailed. Following a similar strategy to the one in *Ball toss* in the *Clear Speech* app (CUP, 2011), one of the activities in the app requires learners to throw a ball into the one of three baskets, depending on the sound they hear. As a case in point, Figure 73 illustrates this visually in an activity aimed at teaching the pronunciation of <-ed> endings. Learners would have to make their choices based on whether they hear the word ending pronounced with /t/, /d/ or /ɪd/. Even though the project is still in its early stages, the researcher intends to continue exploring the potential of mobile apps with the software he is developing.

Figure 73. Sample screenshot of a gamified identification activity from the researcher's app

7.5 Recommendations for practice

Implementations of these types of instruction in FL classes should be free from some of the hindrances encountered here, such as having to keep a strict control of the learners' task-completion or ensuring that they study the materials presented by teachers. In this regard, training with podcasts, for instance, could be implemented entirely online, as in the pilot study in chapter IV.

Teachers interested in using podcasts as part of their course should be careful to include as many exemplars of the target sounds as possible. This could be easily done by transcribing scripts automatically with software such as PhoTransEdit (PhoTransEdit, 2013), the one employed in study 1 (see chapter IV). Perception activities like the ones employed on stage 1 in study 1 could be easily conducted in VLEs such as Moodle or Sakai, offering learners automatic feedback on their responses. Institutional VLEs were not used in study 1 given the need to administer materials only to the experimental group. However, for implementations of this type of instruction that do not need to measure improvements empirically, VLEs may be more convenient than Google Drive questionnaires.

Furthermore, future implementations of this type of podcast-based pronunciation instruction should combine scripted and spontaneous tasks. Ducate and Lomicka (2009) used a combination of these two types of activities and found that learners performed similarly in controlled and extemporaneous production. Nevertheless, the results of study 1 reveal that the learners' improvements in spontaneous production were very moderate, possibly due to the fact that they only had to complete two controlled production tasks. Opportunities for output in study 1 were limited to only two podcasts and these did not foster extemporaneous practice.

Additionally, if learners are required to provide feedback to their peers, it would be convenient to start with simple evaluation sheets that facilitate the learners' assessment, like the multiple-choice activities employed here, and then move on to more open-ended ones where they can offer explanations on how to improve. Open-ended evaluations are problematic because it may be difficult for learners to provide adequate feedback indicating to their peers how to improve. In fact, this was one of the comments in the learners' responses to the questionnaires in the podcasts pilot study. Besides, allowing learners to provide free evaluations may lead some students to focus on aspects other than pronunciation (Lord, 2008). However, for learners to start developing an adequate metalanguage that allows them to reflect on pronunciation and self-monitor their performance (see Couper, 2011), it is vital to have this kind of open-ended assessment. Thus, a possible solution to the above-mentioned difficulties would be to start using multiple-choice items to train learners to identify the problematic aspects, and then, thanks to the explicit instruction they would receive, help learners explain to their classmates how to improve.

With regard to apps, for future implementations to really exploit the potential of this technology, it is imperative that training stimuli includes a much wider range of exemplars of the sounds, both in terms of phonetic contexts and lexical items. As explained in chapter I, learners are supposed to form new phonetic categories after being exposed to numerous instantiations of the target sounds (Flege, 1991). Hence, stimuli should ideally illustrate the target pronunciations as featured by different voices and in different sample words. This would allow learners to become familiarised with the phonological make-up of a wider range of words and foster greater improvements in the learners' controlled and spontaneous production of the target sounds.

Furthermore, given the positive results obtained with the feedback provided with speech-to-text ASR applications (Liakin et al., 2014), once learners have worked on their perception of the target aspects – in an app like the one used in study 2 –, they could test their intelligibility by using this type of speech-to-text apps. In addition, offering learners articulatory information on mispronounced words and phonemes would also be a way of letting them know how to attain those sounds in production (Engwall, 2008).

Finally, regarding Twitter, teachers could send tweets several times in order to help learners remember the tips better. Students in study 3 did not consider reading tweets only once to be enough so as to remember them. Thus, perhaps sending the same tweets on different days as reminders would be a good option. Also, teachers could foster a more active engagement on the part of learners if the latter are required to do something with the tweets. As a case in point, they could ask students what they think may be challenging about a certain word after listening to it and allow them to give the answer, rather than sending explicit information in the tweet from the start as was done in study 3. Asking learners questions about the target aspects may foster a more active participation and may make it easier for learners to consolidate the information. Moreover, interacting actively through Twitter may also increase the learners' motivation and foster opportunities for vicarious learning; for example, when students provide their own examples, comments, or make requests for clarification.

7.6 Concluding remarks

The studies in this dissertation offer empirical evidence that the three tools explored can help learners modify aspects of their pronunciation that are difficult to improve without instruction. Moreover, the data reported here add to the extensive evidence supporting that training learners' perception can lead to improvements in their production (e.g. Bradlow et al., 1997; Lambacher et al., 2005; Wong, 2015). Given the impossibility to provide learners with adequate automatic feedback telling them how to improve their pronunciation, efforts should be made in order to help them create adequate mental targets to guide their subsequent productions and self-monitor their pronunciation. Technology should be used to help learners perceive and produce key aspects of FL phonology adequately. In this way, classroom time could be best employed in order to foster the automatisation of aspects that learners practise on their own in communicative activities with other partners; by moving from an imitative and controlled kind of practice at home, to more spontaneous interactions in the classroom.

Appendices

Appendix 1. Sample of the consent form used for study 2.

Consentimiento Informado para Participantes de Investigación

El propósito de esta ficha de consentimiento es proveer a los participantes en esta investigación con una clara explicación de la naturaleza de la misma, así como de su rol en ella como participantes.

La presente investigación es conducida por Jonás Fouz González, investigador en la Universidad de Murcia y en la UCAM. La meta de este estudio es investigar la eficacia de las aplicaciones móviles como herramienta de aprendizaje para la mejora de las destrezas orales.

Si usted accede a participar en este estudio, se le pedirá que realice un **total de 4 entrevistas** (5 entrevistas si está en el grupo 2) con el investigador y que conteste a dos cuestionarios. De estas entrevistas, dos tendrán lugar al inicio del estudio y dos al final.*

- La primera y tercera entrevista se realizarán de manera individual y serán grabadas en audio (llevarán aproximadamente 10 minutos de su tiempo).
- La segunda y cuarta entrevista se realizarán de manera grupal y no serán grabadas (tendrán una duración aproximada de 35 minutos).**

Debido a que la presente investigación tiene un carácter formativo, además de estas entrevistas, tendrá que utilizar la aplicación propuesta durante unos 20 minutos al día durante las 2 semanas que dura el estudio. Al firmar este documento **se compromete a hacer un uso adecuado de la aplicación y a mantener la confidencialidad que el investigador le solicita. Es imprescindible que no hable del estudio con otros participantes durante la duración del mismo, ya que esto podría sesgar los resultados obtenidos y resultaría en una pérdida considerable de tiempo y esfuerzo del investigador y del resto de participantes.**

La **participación** es este estudio es **estrictamente voluntaria**. Toda la información recogida en el estudio, así como respuestas a cuestionarios serán codificadas usando un número de identificación y por lo tanto, **serán totalmente anónimas**.

Si tiene alguna duda sobre este proyecto, puede hacer preguntas en cualquier momento durante su participación en él. Igualmente, puede retirarse del proyecto en cualquier momento sin que eso le perjudique de ninguna forma. Si alguna de las preguntas durante la entrevista le parecen incómodas, tiene usted el derecho de hacérselo saber al investigador o de no responderlas. De tener preguntas sobre mi participación en este estudio, puedo contactar con Jonás en el correo electrónico j.fouzgonzalez@um.es.

Desde ya le agradezco su participación.

***Compromiso de reembolso de la aplicación:**

La participación en este estudio implica la compra de la aplicación que el investigador recomiende. Esta aplicación tiene un coste aproximado de 5€, pero este importe le será reembolsado al final la investigación como agradecimiento por su participación y podrá seguir utilizando la aplicación una vez finalizado el estudio. IMPORTANTE: el importe de la aplicación se reembolsará únicamente a los participantes que demuestren haber seguido las pautas indicadas por el investigador de manera adecuada y que finalicen el estudio completando las dos encuestas y todas las entrevistas.

Acepto participar voluntariamente en esta investigación, conducida por Jonás Fouz González. He sido informado/a de que la meta de este estudio es investigar la eficacia de las aplicaciones móviles como herramienta de aprendizaje para la mejora de las destrezas orales.

Me han indicado también que tendré que realizar un total de 4 entrevistas (5 si estoy en el grupo 2), dos al inicio y dos al final del estudio, dos de las cuales serán grabadas en audio. Además, me han informado de que se me reembolsará el importe de la aplicación utilizada para el estudio siempre y cuando cumpla las pautas indicadas por el investigador y finalice el estudio.

Reconozco que la información que yo provea en el curso de esta investigación es estrictamente confidencial. He sido informado de que puedo hacer preguntas sobre el proyecto en cualquier momento y que puedo retirarme del mismo cuando así lo decida, sin que esto acarree perjuicio alguno para mi persona. De tener preguntas sobre mi participación en este estudio, puedo contactar con Jonás en el correo electrónico j.fouzgonzalez@um.es.

Entiendo que una copia de esta ficha de consentimiento me será entregada, y que puedo pedir información sobre los resultados de este estudio cuando éste haya concluido.

Nombre del Participante Firma del Participante Fecha

* Si está en el grupo 2, tendrá una entrevista final adicional.

** La 2ª entrevista del principio del estudio llevará unos 45 minutos debido a que se le enseñará a comprar y utilizar la aplicación y se explicará el procedimiento que vamos a seguir.

Appendix 2. Sample of the confidentiality agreement used in study 1

UNIVERSIDAD DE MURCIA

COMPROMISO DE CONFIDENCIALIDAD

D./Dña. ____, con D.N.I. Nº _______________, se compromete a mantener la confidencialidad necesaria para participar adecuadamente en la investigación llevada a cabo por el profesor Jonás Fouz González. Concretamente, se compromete a hacer un uso adecuado de los materiales utilizados y a no comentar con nadie nada en absoluto acerca del contenido de los materiales que maneja en esta investigación ni sobre el funcionamiento de la misma.

Puesto que no va a ser posible atender a todos los participantes al mismo tiempo y uno de los grupos que se creen participará en otro momento, le rogamos no comente con otros compañeros si usted está en un grupo u en otro. De este modo permitirá no revelar información a los futuros participantes, al mismo tiempo que conseguimos un mayor anonimato, ya que aunque las voces de algunos compañeros puedan sonarle familiares, ningún participante sabrá quién está participando ahora y quién participará más adelante.

Asimismo, debido a que la investigación requiere la grabación semanal de podcasts producidos por los participantes, al firmar este documento se compromete a mantener un comportamiento ejemplar, actuar siempre desde el respeto a los compañeros y a <u>no distribuir ni mostrar a otras personas</u> las grabaciones que se manejen en la investigación.

Manifiesta también que es consciente del compromiso que adquiere a la firma de este documento y que el incumplimiento del mismo derivaría en una lamentable pérdida del tiempo del investigador, de los demás participantes y del suyo propio.

Para que conste, firma este documento en la fecha y lugar abajo indicados.

Fdo.: ____

Murcia, a _____ de octubre de 2014.

Appendix 3. Sample prompt with instructions for the podcasts students recorded the pilot study

Podcast 1 – Angry people

Task 1: After listening to the podcast, read the text below and record it.
Task 2: Please include your opinion (or personal experience) about the content of the podcast at the end of the recording. *What do you think, are people angrier than ever? What's the thing that annoys you most? Any interesting experiences you want to tell us about?*

Instructions/recommendations:

a. Please, record the text in only 1 'try'. Don't record sentence by sentence (this would not help you at all). *Por favor, para que esta actividad te ayude a mejorar, graba el texto 'de una vez', no vayas grabando frase por frase, esto perjudicaría mucho tu pronunciación a nivel de frase y no te permitiría progresar.*
b. Pay special attention to this week's goals /p,t,k,b,d,g/ and /v/.
c. Try to use an engaging voice and sound friendly (don't use a 'monotone' /robotic voice).
d. Be calm & enjoy it!

Dialogue:

Rob: My name is Rob and I'm joined in the studio by Jennifer.
Jennifer: Hello.
Rob: Hello. In today's programme we are talking about anger – that's the strong feeling you get when you feel someone has treated you badly or unfairly. Now, does that sound familiar Jen?
Jennifer: Oh yes. There are many things that make me lose my temper – usually just things that annoy me.
Rob: Well, we'll hear what they are soon and we'll look at some research that says our modern life is making us angrier. But first, keep calm Jen and see if you can answer today's question.
Jennifer: It's OK Rob, I'm in a good mood – I feel happy – so let me have it!
Rob: OK. In a BBC survey, what was found to be the thing that made British people most annoyed? Was it:
a) Someone jumping the queue
b) Delays on public transport
c) Being kept on hold by a call centre
Jennifer: I'll go for c) being kept on hold by a call centre because that's very annoying.
Rob: OK, well, we'll find out if you're right at the end of the programme. So Jen, you say that annoys you, does anything else annoy or anger you?
Jennifer: Public transport annoys me but it's the passengers that I find most annoying especially when they push and shove and cram onto a train. How about you Rob?
Rob: Well, for me, it's got to be rudeness. It really makes my blood boil when people who work in shops are rude to me, the customer – it is as if they don't want me to buy anything! But I suppose that is quite small compared with things that used to make us angry.
Jennifer: Yes. Humans developed the feeling of anger as a basic survival skill – the emotion of anger helped us to do things – so hunger would make us angry and that would make us look for food.
Rob: Interesting stuff. But now we start fuming – so we get very angry – by just small things which aren't that important. This is according to new research published by the University of Central Lancashire in the UK.
Jennifer: The research found people today are angrier than ever. And Doctor Sandi Mann from the university says it is modern life that's to blame.
**Here is where you give your opinion on the podcast.*

Appendix 4. Initial questionnaire pilot study (podcasts)

Encuesta inicial

*Obligatorio

Número de participante * __________
Edad* __________
Sexo *
- Hombre
- Mujer

¿A qué grupo del estudio perteneces? ¿Grupo 1 o Grupo 2? *
- Grupo 1 (el que empieza ahora el estudio)
- Grupo 2 (el que empieza el curso que viene)

¿Cuál es tu nacionalidad? *

____

¿Cómo describirías tu nivel de inglés? *
(Los niveles entre paréntesis son los descriptores del Marco Europeo de Referencia).
- Muy alto, bilingüe o casi bilingüe (nivel C2)
- Bastante alto (nivel C1)
- Avanzado (nivel B2)
- Intermedio (nivel B1)
- Elemental (nivel A2)
- Principiante (nivel A1)

¿Tienes algún título oficial que certifique tu nivel de inglés? *
- Si
- No
- Otro: __________

¿Cuántas HORAS has pasado estudiando inglés a lo largo de tu vida? *
Ten en cuenta que en secundaria y bachillerato se estudia inglés una media de 105 horas anuales (3 horas a la semana durante 35 semanas -quitando festivos y vacaciones). Por favor, haz un cálculo (honesto) aproximado. Suma si has asistido a Escuelas de Idiomas, academias, etc..

____

¿Has estado alguna vez en un país de habla inglesa? * (Inglaterra, Gales, Escocia, Irlanda, Estados Unidos, Australia, etc.)
- Sí
- No

De ser afirmativa la pregunta anterior, ¿cuánto tiempo has pasado en un país de habla inglesa? (el total de tus viajes)
- Entre 1 y 2 semanas
- Entre 2 semanas y 1 mes
- Entre 1 y 2 meses
- Más de 2 meses
- Otro: __________

¿Cuánto tiempo estás expuesto al inglés habitualmente? * (En clase, viendo la televisión, escuchando la radio, etc.)
- Menos de 1 hora al día
- Entre 1 y 2 horas al día
- Más de 2 horas al día
- Otro:__________

SIN CONTAR LAS HORAS DE CLASE, ¿cuánto tiempo estás expuesto al inglés en tu tiempo libre? * (Viendo la televisión/cine, escuchando la radio, con videojuegos, con amigos, etc.)

- Menos de 1 hora al día.
- Entre 1 y 2 horas al día.
- Más de 2 horas al día.
- Entre 1 y 2 horas a la semana.
- Otro: __________

¿Cómo? * (Puedes marcar varias opciones).

- Viendo la televisión.
- Escuchando la radio.
- Escuchando música.
- Hablando con amigos.
- Otro: __________

Por favor, ordena en orden de dificultad los aspectos del inglés que más difíciles te resultan (siendo 1 lo más difícil y 4 lo más fácil). *

Cuando la gente te habla; cuando lees; cuando hablas; cuando escribes; (otros).

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Me da mucha vergüenza hablar inglés en público.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Me da vergüenza grabar mi voz y que me oiga otra gente.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Presto atención a cómo la gente pronuncia en inglés y me comparo con ellos (La forma en que acentúan las palabras, cómo pronuncian los sonidos, la entonación que usan, etc.)

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Cuando cometo fallos de pronunciación, me doy cuenta.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Cuando cometo fallos de pronunciación, soy consciente de cuál es el problema.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *

Cuando cometo fallos de pronunciación, soy capaz de corregir el problema yo mismo.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

De ser afirmativa la anterior, ¿cómo? *

- Con un diccionario
- Con la función de 'pronunciar' del traductor de Google o similar
- Le pregunto a alguien
- No hago nada
- Otro: __________

Por favor, reflexiona sobre tu pronunciación (en inglés) y ordena los siguientes aspectos del 1 al 4, siendo 1 el que más trabajo te cuesta y 4 el que menos trabajo te cuesta. *

Pronunciar bien los sonidos; acentuar las palabras correctamente; usar un ritmo adecuado (acento a nivel de frase); utilizar la entonación correcta.

____

¿Cómo de importante es para ti conseguir una pronunciación nativa (que alguien te oiga y no sepa si eres inglés o no)? *

- Muy importante
- Importante
- Relativamente importante, pero me contento con que me entiendan
- Mi meta principal es que la gente me entienda, no me importa pronunciar mal si consigo transmitir el mensaje que quiero transmitir
- Otro: __________

¿Qué modelo de inglés prefieres? *

- Británico.
- Americano.
- Otro: ___________

¿Por qué prefieres uno u otro? *

____

Sobre los podcasts... *

- No sabía lo que eran hasta ahora.
- Sé lo que son, pero no los he utilizado nunca.
- Los he utilizado alguna vez, pero no para grabar mis propios podcasts, sólo para escuchar los que hace otra gente.
- Escucho podcasts habitualmente.
- Escucho podcasts habitualmente e incluso grabo mis propios podcasts a veces.

Appendix 5. Final questionnaire pilot study (podcasts)

Encuesta final

*Obligatorio

Edad * __________

Sexo *

- Hombre
- Mujer

¿Cuánto tiempo pasas al día en internet? *

- 30 minutos o menos
- Entre 30 minutos y 1 hora
- Entre 1 y 2 horas
- Más de 2 horas
- Otro: __________

¿Dónde te sueles conectar a internet? * (Puedes elegir varias opciones)

- Desde el móvil (por la calle)
- Desde el móvil (en casa)
- Desde la universidad (en las alas de informática)
- Desde casa (en un ordenador)
- Otro: __________

Si utilizas las redes sociales, ¿qué redes sociales usas? * (Puedes elegir varias opciones)

- Facebook
- Twitter
- Tuenti
- Badoo
- LinkedIn
- Google+
- Instagram
- Otro: __________

Si utilizas las redes sociales, ¿cuánto tiempo sueles pasar en ellas? *

- 30 minutos al día o menos
- Entre 30 minutos y 1 hora al día
- Entre 1 y 2 horas al día
- Más de 2 horas al día

¿Tienes un smartphone? *

- Sí
- No

De ser afirmativa la pregunta anterior, ¿qué sistema operativo tiene?

- Android
- Windows
- iOS (Apple)
- Blackberry
- Otro: __________

¿Tienes un iPod, mp3 o reproductor de música similar? *

- Sí
- No

Si escuchas música, la radio o similar ¿dónde lo haces habitualmente?
(Puedes elegir varias opciones)

- Móvil
- iPod, mp3 o similar
- Ordenador
- Otro: __________

Si tienes alguna duda sobre el funcionamiento de alguno de los aparatos electrónicos que utilizas, ¿cómo la resuelves? *
(No tiene por qué limitarse a aparatos de música, cualquier aparato electrónico - teléfono móvil, ordenador, tablets, televisor, ebooks, DVD, etc.)

- Lees el manual de instrucciones.
- Intentas averiguar cómo funciona investigando los diferentes menús y opciones.
- Le preguntas a alguien.
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?)
* Me da mucha vergüenza hablar inglés en público

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Me da vergüenza grabar mi voz y que me oiga la gente.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Presto atención a cómo la gente pronuncia en inglés y me comparo con ellos (La forma en que acentúan las palabras, cómo pronuncian los sonidos, la entonación que usan, etc.)

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, me doy cuenta.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy consciente de cuál es el problema.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy capaz de corregir el problema yo mismo.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

Reflexionando sobre tu nivel de implicación en el estudio en el que acabas de participar. ¿Cómo describirías el esfuerzo realizado? *

1. Bajo (me he esforzado poco).
2. Bajo-medio (me he esforzado lo justo).
3. Medio (me he esforzado pero podría haberme esforzado más).
4. Alto (me he esforzado mucho, pero creo que podría haber puesto todavía más empeño).
5. Muy alto (he dado lo mejor de mí mismo).

De los podcasts recibidos, ¿cuál te ha parecido más útil? *

____

¿Leías la información de los PowerPoints antes de grabar el podcast? *

1. Nunca
2. Algunas veces
3. A veces
4. Casi siempre
5. Siempre

¿Cuánto tiempo te ha llevado hacer las tareas del estudio cada semana? *

____

¿Cuántos 'intentos' de cada podcast has grabado? *
Es decir, si no lo grababas a la primera, normalmente, ¿cuántos ensayos realizabas antes de la versión final?

____

¿Te ha parecido muy exigente el estudio (en cuanto al tiempo requerido semanalmente)? *

- Sí
- No
- Otro: __________

¿Crees que la temporalización de 'un podcast por semana' es apropiada? *

- Sí
- No

Si pudieras elegir el tiempo dedicado para cada podcast, ¿cuánto tiempo le darías? *
Si deseas añadir algún comentario a este respecto, utiliza la opción "otro".

- Una semana para cada uno
- Semana y media
- Dos semanas
- Otro: __________

¿Te ha parecido apropiada la duración del estudio (5 podcasts)? *
Si deseas añadir algún comentario a este respecto, utiliza la opción "otro".

- Sí
- No
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? *
El feedback recibido por mis compañeros ha sido adecuado y útil, me ha ayudado a mejorar.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Leías el feedback del profesor y los compañeros antes de grabar tu siguiente podcast? *

- Sí, siempre.
- A veces.
- Sólo el de mis compañeros.
- Sólo el del profesor.
- Casi nunca.
- Nunca
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? *
Considero que el feedback que he proporcionado a mis compañeros ha sido útil y les ha ayudado a mejorar.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Me he sentido capacitado para dar feedback a los compañeros, sabía qué pronunciaban mal y cómo ayudarlos a mejorar.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

Por favor, si quieres comentar algo más con respecto al feedback (cualquiera de los tipos) hazlo aquí.

____

¿Cuál es tu postura sobre la siguiente afirmación? *
El estudio me ha ayudado a ser consciente de mi pronunciación y a darme cuenta de los errores que cometo cuando hablo (he mejorado más que si no hubiera participado en el estudio)
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Creo que podré utilizar lo que he aprendido durante el estudio en mi uso diario del inglés.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Creo que el estudio me ha ayudado a ser más consciente de cómo pronuncia la gente cuando habla en inglés.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

El estudio me ha ayudado a mejorar mi pronunciación... *

- Nada
- Muy poco
- Un poco
- Bastante
- Mucho

Para futuras aplicaciones del mismo estudio. ¿Qué cambiarías o mejorarías? Por favor, piensa qué aspectos podrían mejorar. __________

Appendix 6. Materials used for training in study 1

Input: podcasts learners received

Podcast 1/s/-/z/: ***Is silence golden?*** **(edited version)**

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2013/03/130328_6min_is_silence_golden.shtml

Rob: We're talking about noise today – and looking at some of the words and phrases associated with noise and its opposite: peace and quiet. But, as always, let's start with a question. A new survey in the UK has identified the ten jobs where people are most exposed to noise – noise that can cause serious damage to someone's hearing. Which one of these three jobs has the most exposure to noise?

a) A nightclub worker
b) A classical musician
c) An airport ground staff worker

Finn: I think it's got to be c) the person who works in the airport. Planes are very noisy aren't they Rob?
Rob: That's true, very noisy.

(Mix of sounds from a busy London street)
Finn: So we heard drills, and buses and church bells in there as well, didn't we Rob? It's a real **din** – or bad noise – but people in urban areas all around the world have to live with that sort of noise all the time.
Rob: Yes but I guess they get used to it and it's all part of city life but it does mean it can be difficult to hear yourself think! And I think you'll agree the world is becoming noisier?
Finn: It is. So let's hear from Doctor Stephen Dance who went to discover how bad the noise in London really is. How does he describe the noise for pedestrians?

Dr Stephen Dance. South Bank University:
We're here looking at the London soundscape, and as ever there is a fire engine going by, just as I'm talking. That is just as loud as it would be on a motorway but we're on a side street, so it's quite deafening for the pedestrians.

Finn: It was – he described it as deafening for pedestrians – so, extremely loud and possibly causing deafness. But how would we know a fire engine was on its way to an emergency without such a sound?
Rob: It's a good point. Sometimes a loud noise is needed so it can be heard over other noises. And in other situations we sometimes make more noise to drown out – or cover up – the sounds we don't want to hear. So we turn our music up to drown out the sound of the washing machine for example!

Finn: But of course if everyone turns up their own music the noise becomes even greater. So, Rob, what is the solution?

Finn: Rob, maybe you should join a noise abatement group - these are groups of people who campaign to control levels of noise. They try to restrict planes flying over residential areas at night and encourage people not to disturb their neighbours by playing music too loudly. Maybe the big question really is what is noise? Some people may call a sound just noise whereas others may say it's music to their ears – a beautiful sound.

Podcast 2 /s/-/z/: ***Are you a Winner?*** **(edited version)**

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2014/03/140313_6min_compers.shtml

Rob: Hello I'm Rob. I'm joined today by Neil. Hello Neil.

Neil: Hello Rob.

Rob: Now Neil, I have a question for you – do you think you are a winner?

Neil: You mean someone who has a lot of success in everything?

Rob: Well, not exactly – I'm just talking about competitions – do you have a lot of success or **luck** in winning them?

Neil: Competitions? No, not at all. I don't think I've ever won a competition.

Rob: Ah, bad luck. That means you're not a 'comper'. That's an informal name for someone who takes part in – or enters – competitions on an almost semi-professional basis. They spend a lot of time trying to win something.

Neil: You mean winning prizes – or free gifts.

Rob: I do. And Neil, you could win a prize if you can correctly answer today's question. So, are you ready?

Neil: I'm ready.

Rob: Well, a lottery is one kind of competition where the prize is money. The biggest cash prizes can be won in the USA – but do you know what the biggest ever cash prize to be paid in America is? Is it:

a) $590 million

b) $890 million

c) $1 billion

Neil: Well, things tend to be big in America, so I'm going to go for c) $1 billion dollars.

Rob: I'd like to win that. We'll find out if you are right or wrong later on. So let's talk more about 'compers' – people who regularly take part in competitions. We could say they are hooked on – meaning addicted to – taking part.

Neil: Yes, the lure – or attractiveness – of winning big prizes means these people just can't stop answering quiz questions, writing slogans and captions or solving puzzles.

Rob: Some people go to great lengths – or put a lot of effort into winning something – even if it's just a box of chocolates or a coffee mug. It's just the excitement of winning.

Neil: But sometimes there are big prizes to win – a new car, a speedboat or a holiday of a lifetime. The only problem is that these prizes are either not easy to win or there are millions of people trying to win them.

Rob: I've certainly never won anything as fantastic as that – but one man who has had plenty of good luck is Martin Dove, who is a retired lecturer and an expert 'comper'.

Neil: He certainly is. He's won a yacht, a racehorse and lots of smaller prizes too. Let's hear from him now. Listen out for the names he says people have called him...

Martin Dove, a 'comper': *I've been a comper for 40 years. It's like admitting some addiction isn't it really! Some people have called me the Master of Comping, the King of Comping, the Guru of Comping, but it's just a word, it's just a phrase, it's just I was fairly high-profile.*

Rob: So, he says he was fairly high-profile – that means he was often seen in public, mentioned in newspapers, or appeared on television. And because he was high-profile he got called a few nicknames...

Neil:...names like the master of comping – so someone who is very good or skilled at it. And the king of comping – not an actual royal king but someone is the best at doing something. And the guru of comping – that's someone who other people respect and go to for advice about comping.

Podcast 3 /s/ - /z/: ***Modern offices*** **(edited version)**

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2014/05/140501_6min_modern_offices.shtml

Rob: Hello, I'm Rob. Welcome to 6 Minute English. I've got Finn with me today. Hello Finn.

Finn: Hi Rob, how's it going?

Rob: All right. You might notice it's a bit noisier than usual – that's because we're in our open-plan office.

Finn: Yes, it's a big room full of desks with no walls between them and, as I look around, I can see maybe about 50 colleagues working very hard at their computers!

Rob: And today we're talking about open-plan offices – and learning some language related to office life.

Finn: Rob – a question? You know a lot about sound, don't you?

Rob: Well, a bit.

Finn: What do we call a kind of noise that contains the full range of sounds that humans can hear? Is it... a) white noise b) green noise c) pink noise

Rob: Good question. I'm only familiar with the term 'white noise', so I'll go for a) white noise.

Finn: Well, we'll see if you're right at the end of the programme. So shall we continue talking about offices?

Rob: Yes, millions of people like us work in open-plan offices these days, but they're not new. Do you remember Henry Ford, the American industrialist from the late 19th Century and early 20th Century?

Finn: Yes, he owned factories and he made the famous Ford cars.

Rob: He's also one of the main names in the story of open-plan offices.

Finn: Yes, Henry Ford was really concerned with efficiency, wasn't he?

Rob: Yes. Efficiency is one of main reasons for open-plan offices – they increase communication and collaboration among staff.

Finn: Now, a company's staff – its employees – work together for the same goals – they collaborate, exchanging information and ideas. This can be nice, but there can be too many of us in a small space!

Rob: Franklin Becker, social psychologist at Cornell University in the US, thinks the reason open-plan offices have become acceptable and popular, or as he says – the reason they have taken root – is different. What reason does he give?

Franklin Becker, social psychologist at Cornell University, US: *The fundamental reason* *why open plan has taken root has nothing to do really with communication or collaboration or even flexibility. It has to do with the fact that you can reduce the amount of space per person in an open-plan versus any kind of a closed cellular office.*

Finn: Well, it's all about saving space and money. He says it takes less space per person in an open-plan office than it does in a cellular office – that's an office which is made up of lots of small, closed rooms.

Rob: In those offices, the space for each individual – per person - is limited. Which is why open-plan offices have taken root.

Finn: So some very good reasons for open-plan offices. But what about the noise?

Rob: The noise! Yes! Although open-plan offices can save a company money, they have hidden costs. Sound expert Julian Treasure explains what they are. He uses a very important word for business. Which word is it?

Julian Treasure, chairman of the Sound Agency: *Nobody can understand two people talking at the same time. We have bandwidth for about 1.6 people talking. Now that's key when we are talking about open-plan offices because if I'm trying to do work it requires me to listen to a voice in my head to organise symbols, to organise a flow of words and put them on paper, for example. And if you're talking at the same time, then you're taking up one of my 1.6. I'm left with 0.6 in my head. That doesn't work very well - it reduces my productivity dramatically.*

Podcast 1 /b d g/: *Odd job intervies* (edited version)

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2014/02/140227_6min_odd_interviews.shtml

Rob: Welcome to 6 Minute English with me Rob.

Finn: And me, Finn.

Rob: Now, Finn, I'd like to start by asking you: How many job interviews have you had?

Finn: That's difficult. Maybe ten interviews in my life.

Rob: Ten. That's quite a few – and do you enjoy going to job interviews?

Finn: I absolutely adore them! No, I'm joking! Who does? Nobody does.

Rob: Well, for me, they are torture! I hate being grilled by a panel – or group – of people. I know I can do the job but I hate having to convince them! Today, we'll be discussing some odd job interviews and looking at some related vocabulary. So Finn, are you ready for your first interview question?

Finn: Yes Rob, I am raring to go!

Rob: Good to hear. Well, it's important to know what type of job you are being interviewed for. Some job titles are a bit exaggerated. So, what type of job has been named a 'Field Nourishment Consultant'? Is it: a) A waitress b) A school dinner lady c) A petrol station assistant

Finn: I think that it's b) a school dinner lady.

Rob: An interesting choice. I'll let you know if you are wrong or right later on. Let's talk more about job interviews. A traditional interview usually involves being asked a list of questions, and sometimes you have to give a short presentation.

Finn: Yes, questions like: "Why do you want this job?" or, "Where do you see yourself in five years' time?"

Rob: Yeah, but some interviewers – the people who ask the questions – go a bit further and ask the interviewees – the people being interviewed – to do some inappropriate things.

Finn: You mean they are asked do things are not really relevant to the job. Such as Alan Bacon, a university graduate, who last year was asked to do a dance as part of his interview.

Rob: Well, maybe the position – or job – was for a dancer or a children's entertainer?

Finn: No – it was actually for a job as a sales assistant in an electronics shop; so, someone who works on the shop floor, giving advice to customers about what to buy. There's no dancing involved.

Rob: Well, let's hear from him now.

Alan Bacon, university graduate: *We all wanted the job, some of us are desperate, like myself, and the idea is just to keep smiling and go for it. On the surface I had to look positive, I was smiling, I was laughing along with it, but inside I felt degraded and humiliated especially.*

Finn: Well, later on, he did complain and he got an apology.

Rob: But experts say there are now too many candidates chasing too few jobs so companies are trying unorthodox – non-traditional ways of recruiting people - to see who stands out.

Finn: Yes, well, in any job interview it's good to leave a lasting impression – that means to get noticed and make people remember you. I suppose doing a dance is a good way of breaking the ice – making people feel relaxed – but being asked to do something outside your comfort zone also seems a bit unfair to me.

Rob: Yes, but I guess if you want that job, you'll do anything.

Finn: Well, almost!

Rob: We hope you've enjoyed today's 6 Minute English. Please join us again soon for another programme.

Both: Bye.

Podcast 2 /b d g/: *Learn a thousand foreign words* (edited version)

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2014/02/140206_6min_1000_words.shtml

Rob: Hello I'm Rob. Welcome to 6 Minute English. I'm joined today by Finn. Hi Finn.

Finn: Hi Rob – or should I say 'ni hao' Rob?

Rob: Your Chinese is very good Finn but I wonder how many more Chinese words you know? 1,000 perhaps?

Finn: (In Chinese: "Not really, I just know a little...")

Rob: Now you're just showing off! Not being able to speak a foreign language is a bit of a British trait – or a particular British characteristic. We're not very good at it although Finn is an exception, he can speak many foreign languages, can't you?

Finn: Not that many – a bit of German, some French, Polish a little, Chinese of course, Hokkien, a bit of Japanese... That's about it.

Rob: I'm impressed Finn. Well, now the rest of us Brits are being encouraged to learn at least 1,000 words of another language. We'll talk more about that soon but before I start learning my new words, how about a question Finn?

Finn: Très bien!

Rob: Do you know which is the second most spoken language in England? Is it: a) Polish b) Urdu c) French

Finn: I think I know this one, Rob. I'm going to say a) Polish.

Rob: OK, well, as always I'll let you know the answer at the end of programme. So, as I mentioned, the British are generally considered to be lazy linguists – they just don't bother to learn another language.

Finn: I guess the main reason is that when British people travel around the world they find that English is spoken almost everywhere – so they get by – they survive on just using their native language.

Rob: I think, in the past, the education system was also to blame. Learning a foreign language was not compulsory – it didn't have to be studied - when I went to school, we didn't have to study languages to exam level – so I took the easy option and studied photography instead of French! But of course I regret it now. But that wasn't the same for you Finn?

Finn: At first I didn't really like it but you know, I love words, and then one day I discovered the Chinese language and thought this was fascinating, and it's a key to a whole new culture.

Rob: Well, recently a campaign was launched for those of us who didn't share your enthusiasm or have the opportunity to learn another language. The *1,000 Words* campaign is encouraging everyone in the UK to learn at least 1,000 words of another language. It hopes to help Britain increase international trade.

Finn: The group says that a vocabulary of 1,000 words would allow a speaker to hold a simple conversation. It sounds like a good idea.

Rob: Si! Well, let's hear from the former England footballer and TV presenter, Gary Lineker, who is supporting the campaign. Can you hear what three things he says learning another language gives you?

Gary Lineker: *I think it gives you self-satisfaction and self-esteem if you can speak another language when you're travelling. I think it also gives you an edge in a lot of different areas in the workplace, not just football.*

Finn: So Gary Lineker says there are three things it gives you; it gives you self-satisfaction, firstly. He means you feel good about learning a new skill. But it can have negative meaning – self-satisfaction - that you are smug or pleased with yourself. I don't think he means that here though.

Rob: He also says it improves self-esteem – so you feel good about yourself and it boosts your confidence. Imagine going on holiday to Spain and being able to converse with – or speak to – the locals.

Finn: It feels good! And he also mentioned the economic benefits of speaking another language; it gives you the edge in the workplace. That means it gives you an advantage, especially if you are dealing with foreign companies.

Rob: And it also shows politeness and respect for other people by showing you have made an effort. It's something another footballer, Gareth Bale, has tried to do. Last year he signed for Real Madrid so he tried to master – or to be very good at – speaking Spanish so he could talk to his fans. This is how he got on:

Gareth Bale: *(In Spanish: Hello. It is a dream to play for Real Madrid. Thank you.)*

Rob: Muy bien! Impressive – I think he was saying it was his dream to play for Real

Madrid.

Finn: Rob, your Spanish is very impressive too there. For Gareth Bale, speaking Spanish will help him fit in – perhaps make him more accepted by his teammates and his fans.

Podcast 3 /b d g/: *Young, British and sober* (edited version)

http://www.bbc.co.uk/worldservice/learningenglish/general/sixminute/2014/04/140424_6min_sober.shtml

Rob: Today we're talking about the increase in the number of young people who have decided not to drink alcohol.

Finn: We're particularly talking about young people in the UK – teenagers and people in their early twenties. In some countries, drinking is not a big issue, but in Britain, we're seeing a change in attitude – in other words, a change in people's beliefs and behaviour.

Rob: Now, I started to drink alcohol in my mid-teens. It made me feel grown- up and helped me let my hair down at parties. It didn't always taste good though! What about you Neil – when did you start drinking?

Finn: Just like you Rob, I started drinking in my mid-teens, I'd say. It's a big part of the British social life I would say.

Rob: Well, teenagers in the UK may have had a reputation – they were known for binge drinking. That means drinking lots of alcohol and very quickly. But as we'll hear, that image is changing. And we'll be looking at some vocabulary around the topic of drinking.

Rob: OK, well, alcoholic drinks can be measured in units. This gives drinkers a guide to how much they've consumed – or drunk. Drinking too many units of alcohol can be bad for your health. So how well do you know your units Neil? Do you know how many units there are in a typical bottle of wine?

Finn: I am going to go for 13. I'm pretty confident about this one.

Rob: Unlucky for some, but we'll find out if you are right or wrong later on. Did you know the UK National Health Service recommends that a man should not drink more than 3 or 4 units of alcohol a day, and for women it's 2 or 3 units?

Neil: But we know some young people in the UK who drink far more than that and get intoxicated – another word for getting drunk. And this often leads to a hangover – feeling tired and ill the day after drinking – or worse.

Rob: Well, a report earlier this year by the BBC's Asian Network, found this picture is changing. Young people are drinking less, or not at all – they are abstaining.

Neil: That's good news. I find, even at my age, cutting back on drinking is a hard thing to do. It takes willpower – that's the ability to control my own behaviour.

Rob: Yes, of course this is not an issue for people from certain cultures, who don't drink alcohol because of their religious beliefs. But for others, there are a number of changes that have encouraged younger people to remain sober – or not get drunk.

Neil: Well, there have been awareness campaigns – that's when organised publicity in the media has shown the benefits of not drinking. Also, British pubs and clubs have been stricter when stopping underage drinkers from buying alcohol.

Rob: Yes, because officially you have to be 18 years old before you can buy alcohol. Another reason is the recession – people can't afford to go out and drink.

Neil: But Jonathan Birdwell, who's a senior researcher at the think-tank Demos, has another reason. A think-tank, by the way, is a group of experts brought together, usually by a government, to develop ideas on a particular subject and to make suggestions for action. See if you can hear what his reason is...

Jonathan Birdwell, Senior Researcher for Demos:

Around 2004 we see the rise of awareness campaigns around units to consume, and daily guidelines, we see the arrival of 'drink aware' labels on alcohol. We also see the rise of negative media stories around binge drinking culture.

I think also significant, is the rise of social media technologies, smartphones, iPads. You know we have not only new ways of interacting with our friends, which takes up time, but we also have multiple forms of entertainment that didn't exist, say ten years ago.

Rob: Interesting! One factor for the change is the rise – the increase – in people using social media technology. He calls this significant – so it's important.

Finn: Yes, all this interaction with our friends takes up time. And time is also taken up using what he calls multiple forms of entertainment – things like gaming and watching films online. There's more of it now.

Rob: So maybe one way to cut down on drinking is to keep busy! But does this mean going out boozing – or drinking alcohol – is a thing of the past?

Finn: Not yet, Rob. Although there are more coffee shops and ice cream parlours now, to spend time in – alcohol abuse still remains a problem among the young in the UK.

Rob: Well, on that sobering thought, let me reveal the answer to the question I set you earlier. I asked if you knew how many units there are in a typical bottle of wine.

Finn: I said 13 but I want to change my mind and say b) 10.

Rob: Well, actually, that's a good idea because it is 10. Did you know that one unit equals 10 ml or 8 grams of pure alcohol, which is around the amount of alcohol the average adult can process in one hour? OK, before we go, please could you remind us of some of the words and phrases that we've heard today?

Appendix 7. Training stimuli for study 1 (/s/, /z/, /b/, /d/, /g/), number (nº) of instantiations of the same word during training, and total number of exemplars for each sound

/s/	nº	/z/	nº	/b/	nº	/d/	nº	/g/	nº
so	14	is	25	but	21	and	22	good	13
it's	12	noise	24	job	15	do	14	language	12
offices	11	offices	11	about	11	drinking	14	give	9
just	10	as	9	by	9	good	11	English	5
space	10	has	7	be	8	words	8	go	5
question	7	prizes	6	being	8	drink	7	going	5
some	7	was	6	Rob	8	didn't	6	get	4
office	6	these	5	bit	6	today	5	group	3
success	6	competitions	4	British	6	around	4	programme	3
that's	6	does	4	b)	4	did	4	again	2
C)	5	dollars	4	because	3	media	4	England	2
six	5	music	4	been	3	day	3	Gareth	2
someone	5	reason	4	maybe	3	does	3	Gary	2
sound	5	says	4	vocabulary	3	don't	3	guess	2
yes	5	because	3	able	2	drunk	3	negative	2
pedestrians	3	means	3	Bale	2	find	3	significant	2
person	3	names	3	BBC	2	had	3	smug	2
same	3	pedestrians	3	before	2	idea	3	again	1
say	3	prize	3	behaviour	2	Madrid	3	ago	1
something	3	sometimes	3	beliefs	2	dance	2	big	1
sometimes	3	areas	2	benefits	2	dinner	2	degraded	1
staff	3	closed	2	big	2	down	2	exaggerated	1
biggest	2	example	2	binge	2	drinkers	2	exam	1
Century	2	He's	2	bottle	2	friends	2	gaming	1
certainly	2	jobs	2	Britain	2	I'd	2	getting	1
efficiency	2	millions	2	buy	2	lady	2	giving	1
expert	2	noises	2	bye	2	made	2	glass	1
its	2	noisier	2	footballer	2	mentioned	2	got	1
let's	2	noisy	2	number	2	mid	2	government	1
listen	2	organise	2	ability	1	should	2	graduate	1
lots	2	planes	2	absolutely	1	started	2	grams	1
Master	2	puzzles	2	abstaining	1	studied	2	grilled	1
see	2	reasons	2	abuse	1	traditional	2	grown	1
small	2	sounds	2	back	1	tried	2	guide	1
sounds	2	words	2	Bacon	1	would	2	guidelines	1
takes	2	phrases	1	bad	1	accepted	1	linguists	1
us	2	always	1	Birdwell	1	adore	1	regret	1
versus	2	becomes	1	blame	1	adult	1	together	1
advice	1	bells	1	boosts	1	advantage	1	younger	1
almost	1	buses	1	boozing	1	advice	1		
also	1	business	1	bother	1	afford	1		
answer	1	captions	1	breaking	1	attitude	1		
answering	1	cars	1	Brits	1	candidates	1		

a**ss**o**c**iated	1	cau**se**	1	**b**rought	1	chil**d**ren's	1
ba**s**i**s**	1	cau**s**ing	1	**b**usy	1	confi**d**ence	1
be**s**t	1	colleague**s**	1	**b**uying	1	confi**d**ent	1
bo**x**	1	company'**s**	1	clu**b**s	1	consi**d**ere**d**	1
bu**s**es	1	comper**s**	1	foot**b**all	1	consume**d**	1
busine**ss**	1	computer**s**	1	jo**b**s	1	coul**d**	1
cellular	1	contain**s**	1	la**b**els	1	**d**aily	1
chocolate**s**	1	day**s**	1	no**b**ody	1	**d**ancer	1
city	1	doe**s**n't	1	pro**b**lem	1	**d**ancing	1
cla**ss**ical	1	drill**s**	1	pu**b**licity	1	**d**ealing	1
con**c**erned	1	ear**s**	1	pu**b**s	1	**d**eci**d**e**d**	1
co**s**t**s**	1	ea**s**y	1	remem**b**er	1	**d**egra**d**e**d**	1
cour**se**	1	employee**s**	1	so**b**er	1	**D**emos	1
dan**ce**	1	enter**s**	1	so**b**ering	1	**d**esperate	1
deafne**ss**	1	e**x**actly	1	su**b**ject	1	**d**evelop	1
de**s**cribe	1	explain**s**	1			**d**ifferent	1
de**s**cribed	1	factorie**s**	1			**d**ifficult	1
de**s**k**s**	1	goal**s**	1			**d**iscovere**d**	1
di**s**cover	1	how'**s**	1			**d**iscussing	1
di**s**turb	1	human**s**	1			**d**oing	1
e**x**changing	1	idea**s**	1			**d**ream	1
e**xc**itement	1	i**s**n't	1			**d**rinks	1
e**x**plains	1	level**s**	1			encourage**d**	1
e**x**posed	1	neighbour**s**	1			en**d**	1
e**x**posure	1	newspaper**s**	1			enjoye**d**	1
e**x**tremely	1	nickname**s**	1			fiel**d**	1
fanta**s**tic	1	oppo**s**ite	1			foun**d**	1
fle**x**ibility	1	other**s**	1			grille**d**	1
gift**s**	1	phra**se**	1			gui**d**elines	1
group**s**	1	phra**ses**	1			holi**d**ay	1
gue**ss**	1	question**s**	1			humiliate**d**	1
increa**se**	1	qui**z**	1			i**d**eas	1
indu**s**trialist	1	reduce**s**	1			insi**d**e	1
length**s**	1	require**s**	1			instea**d**	1
le**ss**	1	re**s**idential	1			interviewe**d**	1
mo**s**t	1	room**s**	1			intoxicate**d**	1
mu**s**ician	1	situation**s**	1			iPa**d**s	1
ni**ce**	1	slogan**s**	1			joine**d**	1
noti**ce**	1	someone'**s**	1			lea**d**s	1
pea**ce**	1	symbol**s**	1			measure**d**	1
po**ss**ibly	1	thing**s**	1			min**d**	1
racehor**se**	1	tho**se**	1			name**d**	1
redu**ce**	1	today'**s**	1			nobo**d**y	1
redu**ce**s	1	turn**s**	1			o**dd**	1
save	1	u**s**e**s**	1			ol**d**	1
saving	1	wall**s**	1			organise**d**	1

seen	1	wa**s**n't	1			outsi**de**	1		
semi-	1	wherea**s**	1			rea**d**y	1		
serious	1	year**s**	1			recommen**d**s	1		
side	1					sai**d**	1		
situations	1					secon**d**	1		
skilled	1					signe**d**	1		
slogans	1					spen**d**	1		
smaller	1					stan**d**s	1		
social	1					stu**d**y	1		
solution	1					tire**d**	1		
solving	1					to**d**ay's	1		
sort	1					tra**d**e	1		
soundscape	1					un**d**erage	1		
spend	1					unortho**d**ox	1		
start	1					Ur**d**u	1		
stop	1					won**d**er	1		
story	1					wor**d**	1		
street	1					worl**d**	1		
such	1								
survey	1								
symbols	1								
think**s**	1								
under**s**tand	1								
U**S**	1								
U**S**A	1								
u**s**ed	1								
voi**ce**	1								
work**s**	1								
your**s**elf	1								
ac**c**eptable	1								
119		92		58		107		39	

Appendix 8. Stimuli learners had to read for the output stage in study 1 (including words featuring target sounds in different positions, the texts learners had to read, and the number and percentage of instantiations of each target sound for each text)

Group 1: /s/ - /z/

/s/	/z/	/s/	/z/
Podcast 1		**Podcast 2**	
so (x2)	amazing	some	prizes
sound	musician	spend	easy
silence	residential	some	Brazil
see	reason	spent	prizes
person	visit	six	exactly
just	example	seven (x2)	is (x2)
hates	doesn't	next	there's
peace	music	just	as (x2)
hates	buses	answer	dollars
it's (x2)	is (x2)	question	millionaires
house	noise	casino	was
that's	does	buses	dollars
us	because	that's	lose
	lives		cars
	years		pairs
	was		shoes
	his (x2)		
	enjoys		
	avoids		
	bars		
	clubs		
	cars		

Text 1

/s/ (n = 16; 40%) /z/ (n = 24; 60%) Word count: 81

He *is* the person who hates noise. For example, he doesn't like music. He just enjoys peace and silence, he avoids any sound. He hates bars, clubs, and even cars! It's amazing…
His brother is a musician so… he left his house the moment he was 18 years old. That's the reason why he lives in a residential area. It's so annoying for us… because we never see him now. He could visit from time to time, but he never does.

Text 2

/s/ (n = 15; 40.5%) /z/ (n = 22; 59.5%) Word count: 96

Do you think winning prizes is easy?
You don't need to answer the question. It *is*! There's a casino Brazil where you can win prizes as big as 3 million dollars! Some people go there and just spend two dollars. Next thing you know… they're millionaires! Some guy named Mike, I think it was…He won five million dollars and he only spent six! You have nothing to lose… That's exactly what I need! I would never take buses again… I'd buy seven cars, one for each day. It will be like having seven pairs of shoes.

Group 2: /b, d, g/

	Podcast 1	
/b/	/d/	/g/
been	**d**on't	**g**o
bad	**d**ays	**g**overnment
benefits	**d**ifficult	**g**etting
but	**d**efinitely	**g**ood
be	to**d**ay	**g**ave
may**b**e	a**d**vice	ne**g**ative
a**b**out	nee**d**	a**g**o
jo**b**	ba**d**	hu**g**
	Podcast 2	
/b/	/d/	/g/
BBC	**d**idn't	**g**reat
bit	**d**ialogue	**g**uess
boring	**d**on't	**g**oing
but	**d**ifficult	**g**ot
boy	**d**ifferent	re**g**ret
begging	**d**idn't	pro**g**ramme
a**b**out	yester**d**ay	to**g**ether
a**b**ility	towar**d**s	e**x**ams (x2)
	stu**d**y	be**gg**ing
	confi**d**ent	dialo**gue**
	stu**d**ying	
	love**d**	
	presente**d**	
	improve**d**	
	sai**d**	

Text 1

/b/ (n = 8; 33.3%) /d/ (n = 8; 33.3%) /g/ (n = 8; 33.3%) Word count: 81

May**b**e I **d**on't nee**d** to **g**o to the interview. I've **b**een thinking a**b**out my jo**b** to**d**ay. I was very ne**g**ative when I saw you two **d**ays a**g**o. It's not too **b**a**d**. You see… with the **g**overnment now in power it's very **d**ifficult to see any **b**enefits. The crisis is **g**etting worse. **B**ut thank you for all the **g**ood a**d**vice you **g**ave me. I'll **d**efinitely remem**b**er it and I'll try to **b**e more positive. I'm so happy I want to hu**g** you!

Text 2

/b/ (n = 9; 25.7%) /d/ (n = 15; 42.9%) /g/ (n = 11; 31.4%) Word count: 95

I wonder why you **did**n't like the pro**g**ramme on the **BB**C yester**d**ay. The **d**ialo**gue** was **g**reat… And I love**d** the way they presente**d** it! Although it was a **b**it **b**oring at first… I **d**on't re**g**ret watching it. I **g**uess it improve**d** towar**d**s the end, when **g**oing a**b**out people's a**b**ility to stu**d**y languages. They sai**d** they felt more confi**d**ent when they talk to**g**ether. **B**ut that it **g**ot very **d**ifficult when talking to **d**ifferent people. Also, they **did**n't like stu**d**ying for e**x**ams. When I was a little **b**oy, I kept **b**e**gg**ing my teachers to not have e**x**ams.

Appendix 9. Identification task for the assessment of /b d g/ as stops (study 1)

Todas estas palabras son palabras **españolas**. Escucha la versión inglesa de estas palabras por los auriculares y di si el sonido subrayado se pronuncia igual en los dos idiomas (en términos de **OCLUSIÓN**, es decir si los articuladores están igual de cerrados en los dos idiomas al producir el sonido). **NO** prestes atención a aspectos como la aspiración, la fuerza del sonido, etc.

1. **p**an
2. ri**g**or
3. **b**ar
4. **p**in
5. la**b**or
6. me**g**a
7. referén**d**um
8. **t**ipo
9. **b**anda
10. fi**g**ura
11. cré**d**ito
12. sím**b**olo
13. **b**ate
14. e**c**o
15. **d**ebate
16. **t**é
17. **d**aga
18. **g**ol
19. **b**ase
20. i**d**ioma
21. apén**d**ice
22. toma**t**e
23. **g**orila
24. **b**anana
25. lí**d**er
26. da**g**a
27. a**b**adía
28. in**d**ustria
29. le**g**al
30. **d**ólar
31. o**b**eso
32. **d**anza
33. I**b**eria
34. len**g**ua
35. **c**lan
36. distin**g**ue
37. **d**efensa
38. a**p**arte
39. com**b**inar
40. é**b**ola
41. a**d**icto
42. **g**as
43. me**d**io
44. sin**g**ular
45. **g**uía
46. i**P**od
47. **g**uitarra
48. miem**b**ro

Appendix 10. Stimuli for the sentence-reading task, including distractors (study 1)

	/z/		
	familiar		**novel**
1	The music is in the city.	**9**	The zebra is in the zoo.
2	The dollars are in the office.	**10**	The zombie is in the museum.
3	The shoes are in the shower.	**11**	The *letters* are in the *boxes*.
4	The reason is in the residential.	**12**	The efficiency is in the *cars*.
5	The noise is in the offices.	**13**	The sugar is in the *houses*
6	The advice is easy.	**14**	The cousin is in prison.
7	The example is in the words.	**15**	The *choices* are in the century.
8	The prizes are in the ship.	**16**	The sheep are busy.
		17	The *books* are in the zone.
		18	The zoom is in the *books*.
		19	The zipper is in the *clocks*
	/b d g/		
	familiar		**novel**
20	The benefits are in the study.	**35**	The *book* is in the corridor.
21	The idea is very difficult.	**36**	The *boy* is in the lobby.
22	The programme is on the BBC.	**37**	The rabbit is in the *game*.
23	Is it good? It has to be.	**38**	The tiger is in the *gallery*.
24	The ability is in the guide	**39**	The addict is in the *bars*.
25	The abuse is in the government.	**40**	The *doctor* is with a *girl*.
26	The guess is in the 'maybe'.	**41**	The *bullets* are not legal.
27	The dancer is in the programme.	**42**	The *design* is very modern.
28	I adore the day.	**43**	A *God* is there with the abbot.
29	Did I drink the bottle? No, I didn't.	**44**	The *danger* is the *guard*.
30	Nobody here lives together.	**45**	The sugar is in the yogurt.
31	I'm confident about today.	**46**	The melody is in the *bedrooms*.
32	What's all this about? Did you drink again?	**47**	A global warming will begin.
33	You are a bit negative. I don't think so.	**48**	The *daisy* will adapt.
34	Where is he going? He was there a day ago.	**49**	The hole above is very *deep*.

Note: Words in italics were originally included as novel stimuli in order to offer a greater sample. The novel stimuli for the /s – z/ contrast were initially included in order to test the learners' pronunciation of the /s – z/ distinction in plurals. In stimuli for /b d g/, 15 words featured these sounds in word-initial position flanked by vowels (VCV), including five tokens per sound. However, they were omitted from the analysis due to limitations of time.

Appendix 11. Initial questionnaire for study 1

Encuesta inicial

*Obligatorio

Sexo *

- Hombre
- Mujer

Edad *

____

¿A qué grupo del estudio perteneces? ¿Grupo 1 o Grupo 2? * (Te lo he indicado por email)

- Grupo 1
- Grupo 2

¿Cuánto tiempo pasas al día en internet? *

- 30 minutos o menos
- Entre 30 minutos y 1 hora
- Entre 1 y 2 horas
- Más de 2 horas

¿Dónde te sueles conectar a internet? * (Puedes elegir varias opciones)

- Desde el móvil (por la calle)
- Desde el móvil (en casa)
- Desde la universidad (en las alas de informática)
- Desde casa (en un ordenador)

Si utilizas las redes sociales, ¿qué redes sociales usas? * (Puedes elegir varias opciones)

- Facebook
- Twitter
- Tuenti
- LinkedIn
- Google+
- Instagram

¿Tienes un smartphone? *

- Sí
- No

De ser afirmativa la pregunta anterior, ¿qué sistema operativo tiene?

- Android
- Windows
- iOS (Apple)
- Blackberry
- Otro: ___________

Por favor, indica el modelo de tu teléfono móvil y marca *

____

Si escuchas música, la radio o similar ¿dónde lo haces habitualmente? * (Puedes elegir varias opciones)

- Móvil
- iPod, mp3 o similar
- Ordenador
- Otro: ___________

Reflexiona sobre tu nivel de inglés ¿Cómo lo describirías? *
(Los niveles entre paréntesis son los descriptores del Marco Europeo de Referencia).
- Muy alto, bilingüe o casi bilingüe (nivel C2)
- Bastante alto (nivel C1)
- Avanzado (nivel B2)
- Intermedio (nivel B1)
- Elemental (nivel A2)
- Principiante (nivel A1)

¿Tienes algún título oficial que certifique el nivel de inglés que mencionas arriba? *
- Si
- No
- Otro: __________

¿Has estado alguna vez en un país de habla inglesa? *
(Inglaterra, Gales, Escocia, Irlanda, Estados Unidos, Australia, etc.)
- Sí
- No

De ser afirmativa la pregunta anterior, ¿cuánto tiempo has pasado en un país de habla inglesa?
(el total de tus viajes)
- Entre 1 y 2 semanas
- Entre 2 semanas y 1 mes
- Entre 1 y 2 meses
- Más de 2 meses
- Otro: __________

¿Cuánto tiempo estás expuesto al inglés habitualmente? *
(En clase, viendo la televisión, escuchando la radio, etc.)
- Menos de 1 hora al día
- Entre 1 y 2 horas al día
- Más de 2 horas al día
- Otro: __________

SIN CONTAR LAS HORAS DE CLASE, ¿cuánto tiempo estás expuesto al inglés en tu tiempo libre? *
(Viendo la televisión/cine, escuchando la radio, con videojuegos, con amigos, etc.)
- Menos de 1 hora al día.
- Entre 1 y 2 horas al día.
- Más de 2 horas al día.
- Entre 1 y 2 horas a la semana.
- Otro: __________

¿Cómo? * (Puedes marcar varias opciones)
- Viendo la televisión.
- Escuchando la radio.
- Escuchando música.
- Hablando con amigos.
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? *
Me da mucha vergüenza hablar inglés en público.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Me da vergüenza grabar mi voz y que me oiga otra gente.

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Presto atención a cómo la gente pronuncia en inglés y me comparo con ellos (La forma en que acentúan las palabras, cómo pronuncian los sonidos, la entonación que usan, etc.)
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo
¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, me doy cuenta.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo
¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy consciente de cuál es el problema.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo
¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy capaz de corregir el problema yo mismo.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo
De ser afirmativa la anterior, ¿cómo? *
- Con un diccionario
- Con la función de 'pronunciar' del traductor de Google o similar
- Le pregunto a alguien
- No hago nada
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? *
Para mí, conseguir una pronunciación nativa que alguien me oiga y no sepa si soy inglés o no) es: *
- Muy importante
- Importante
- Relativamente importante, pero me contento con que me entiendan
- Mi meta principal es que la gente me entienda, no me importa pronunciar mal si consigo transmitir el mensaje que quiero transmitir
- Otro: __________

¿Qué modelo de inglés prefieres? *
- Británico.
- Americano.
- Otro: __________

¿Por qué prefieres uno u otro? *

____

Sobre los podcasts... *
- No sabía lo que eran hasta ahora.
- Sé lo que son, pero no los he utilizado nunca.
- Los he utilizado alguna vez, pero no para grabar mis propios podcasts, sólo para escuchar los que hace otra gente.
- Escucho podcasts habitualmente.
- Escucho podcasts habitualmente e incluso grabo mis propios podcasts a veces.

Appendix 12. Final questionnaire for study 1

Encuesta final

***Obligatorio**

¿A qué grupo del estudio perteneces? *

- Grupo 1
- Grupo 2

¿Cuántas HORAS has pasado estudiando inglés a lo largo de tu vida? *

Ten en cuenta que en secundaria y bachillerato se estudia inglés una media de 105 horas anuales (3 horas a la semana durante 35 semanas -quitando festivos y vacaciones). Por favor, haz un cálculo (honesto) aproximado. Suma si has asistido a Escuelas de Idiomas, academias, etc...

____

¿Cómo describirías tu nivel de esfuerzo realizado en el estudio? *

(Por favor, contesta con objetividad)

- Muy bajo (no me he esforzado a penas nada)
- Bajo (me he esforzado lo justo)
- Normal (me he esforzado, pero podría haberme esforzado aún más)
- Alto (he dado lo mejor de mi mismo)

¿Cómo has grabado los podcasts semanales? *

(Por favor, si tu opción no está aquí añádela)

- Desde el móvil
- Desde el ordenador
- Otro: __________

¿Con qué aplicación o programa has grabado los podcasts semanales? *

(Ejemplo: con la grabadora de windows, audacity, aplicación móvil que se llama ...)

____

¿Cual es tu postura sobre la siguiente afirmación? *

Creo que el método utilizado en el curso es útil para mejorar aspectos problemáticos de mi pronunciación

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *

Considero que la duración del curso (3 semanas) ha sido apropiada para mejorar los sonidos problemáticos estudiados

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *

No habría hecho falta tanto tiempo para esos sonidos, con menos habría bastado

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *

Aún necesitaría más práctica para dominar esos sonidos, no ha sido suficiente tiempo

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *

Creo que el método utilizado en el curso me ha ayudado a PERCIBIR (oír) ciertas peculiaridades de los sonidos estudiados

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *

Creo que el método utilizado en el curso me ha ayudado a PRONUNCIAR mejor los sonidos estudiados

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Has mirado el feedback que te daban tus compañeros sobre tus grabaciones para comprobar qué fallos tenías? *

- Siempre
- A veces
- Nunca

¿Cómo definirías el nivel de esfuerzo de los compañeros en la grabación de sus podcasts que has evaluado? *

- Alto (se nota que intentaban dar lo mejor de si mismos y pronunciar bien los sonidos estudiados)
- Normal (grababan la tarea de manera satisfactoria pero no se notaba especial interés)
- Bajo (no se esforzaban en pronunciar bien los sonidos)

¿Cual es tu postura sobre la siguiente afirmación? *
Las encuestas semanales que hacíamos en Edmodo (las actividades 1 y 2 semanales) me han resultado útiles para percibir mejor los aspectos estudiados
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *
El feedback de los compañeros ha sido adecuado, sus calificaciones han sido acertadas
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *
El feedback de los compañeros me ha ayudado a mejorar
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *
Me he sentido capacitado para dar feedback a mis compañeros sobre su pronunciación de los aspectos estudiados
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Qué actividad te ha parecido más útil? *

- Actividad 1 (escuchar el podcast y buscar palabras que contengan el sonido estudiado)
- Actividad 2 (escuchar ejemplos concretos de sonidos muy parecidos y decidir cuál oigo)
- Actividad 3 (GRABAR EN CASA el podcast con los sonidos problemáticos marcados en negrita)

¿Cuántos intentos grababas de cada podcast antes de subir la versión final a Edmodo?*

____

¿Cual es tu postura sobre la siguiente afirmación? *
El estudio me ha parecido muy exigente en cuanto a tiempo, me ha llevado demasiado tiempo
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *
El estudio me ha parecido muy exigente en cuanto a tiempo, me ha llevado demasiado tiempo
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cual es tu postura sobre la siguiente afirmación? *
Creo que los aspectos tratados en el estudio me resultarán útiles para mi uso diario del inglés
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Me da mucha vergüenza hablar inglés en público.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? *
Me da vergüenza grabar mi voz y que me oiga otra gente.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Presto atención a cómo la gente pronuncia en inglés y me comparo con ellos (La forma en que acentúan las palabras, cómo pronuncian los sonidos, la entonación que usan, etc.)
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, me doy cuenta.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy consciente de cuál es el problema.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy capaz de corregir el problema yo mismo.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

Si tienes una duda sobre el funcionamiento de alguno de los aparatos electrónicos que utilizas, ¿cómo la resuelves? *

- Lees el manual de instrucciones
- Intentas averiguar cómo funciona investigando los diferentes menús y opciones
- Le preguntas a alguien
- Otro: __________

Sugerencias/comentarios:
Por favor, escribe aquí cualquier comentario o sugerencia que tengas para ayudarnos a mejorar
____

Appendix 13. Acknowledgement of payment form used in study 2

Al firmar el presente documento certifico que D. Jonás Fouz González, investigador en la UCAM y en la Universidad de Murcia, me ha realizado el reembolso de la aplicación *English File Pronunciation* app por un valor de 5,49€.

Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	
Surname, Name	Fdo. en Murcia, a …….. de noviembre de 2014	

Appendix 14. Instructions on how to use the app in study 2

Procedimiento:
¿Cómo vamos a usar la app?

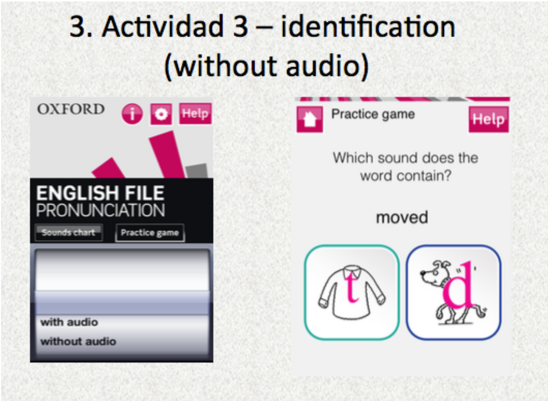
Duración

- 2 semanas de lunes a viernes (5 días a la semana) –si alguien quiere fin de semana, puede.
- +/- 20 minutos diarios (calentamiento + 10 niveles de cada actividad)
- Empezamos **el lunes día 27**.
- Enviaré una encuesta esta semana para que la completéis antes de empezar.

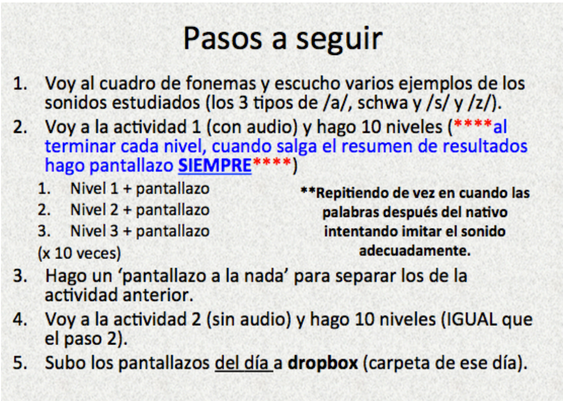
*****IMPORTANTE: estad localizables (correo/ edmodo…)**

Actividades

1. Práctica con el cuadro de fonemas
2. Actividad 1 – sound identification (con sonido) + **pantallazos** con registro de progreso (y miro mis fallos para mejorar)
3. Actividad 2 – sound recall (sin sonido) + **pantallazos** con registro de progreso (y miro mis fallos para mejorar)


3. Actividad 3 – identification (without audio)
OXFORD
Help
ENGLISH FILE
PRONUNCIATION
Sounds chart
Practice game
with audio
without audio
Practice game
Help
Which sound does the word contain?
moved
t
d


Registro de resultados
IMPORTANTE: SIEMPRE que terminemos un nivel, hacemos pantallazo.
Practice game
Help
You scored 5 / 10
moved
secretary
comfortable
without
kitchen
look
doctor
neighbour
friend
boring
New game


Pasos a seguir
1. Voy al cuadro de fonemas y escucho varios ejemplos de los sonidos estudiados (los 3 tipos de /a/, schwa y /s/ y /z/).
2. Voy a la actividad 1 (con audio) y hago 10 niveles (****al terminar cada nivel, cuando salga el resumen de resultados hago pantallazo SIEMPRE****)
1. Nivel 1 + pantallazo
2. Nivel 2 + pantallazo
3. Nivel 3 + pantallazo
(x 10 veces)
**Repitiendo de vez en cuando las palabras después del nativo intentando imitar el sonido adecuadamente.
3. Hago un 'pantallazo a la nada' para separar los de la actividad anterior.
4. Voy a la actividad 2 (sin audio) y hago 10 niveles (IGUAL que el paso 2).
5. Subo los pantallazos del día a dropbox (carpeta de ese día).


Herramientas extra
• Dropbox (para subir los pantallazos y compartirlos conmigo a j.fouzgonzalez@um.es)
• Edmodo (para posibles dudas que puedan surgir)


¿Dudas?

Appendix 15. Percentages of appearance of each sound in the EFP app (study 2)

	Activity 1		**Activity 2**	
Sound	**n = 1000**	**%**	**n = 500**	**%**
/ɪ/	37	3,7	21	4,2
/iː/	23	2,3	18	3,6
/æ/	**68**	**6,8**	**31**	**6,2**
/ɑː/	**34**	**3,4**	**19**	**3,8**
/ʌ/	**32**	**3,2**	**18**	**3,6**
/ə/	**88**	**8,8**	**52**	**10,4**
/ɜː/	40	4	14	2,8
/ɒ/	34	3,4	13	2,6
/ɔː/	41	4,1	17	3,4
/e/	31	3,1	20	4
/ʊ/	69	6,9	28	5,6
/uː/	27	2,7	17	3,4
/eɪ/	31	3,1	15	3
/əʊ/	34	3,4	21	4,2
/aɪ/	37	3,7	15	3
/aʊ/	38	3,8	10	2
/ɔɪ/	0	0	0	0
/ɪə/	24	2,4	19	3,8
/eə/	40	4	10	2
/ʊə/	0	0	0	0
/p/	0	0	0	0
/t/	36	3,6	17	3,4
/k/	0	0	0	0
/b/	0	0	0	0
/d/	32	3,2	11	2,2
/g/	0	0	0	0
/s/	**38**	**3,8**	**21**	**4,2**
/z/	**33**	**3,3**	**19**	**3,8**
/f/	0	0	0	0
/v/	0	0	0	0
/ʃ/	40	4	14	2,8
/ʒ/	0	0	0	0
/θ/	33	3,3	22	4,4
/ð/	33	3,3	15	3
/tʃ/	27	2,7	12	2,4
/dʒ/	0	0	0	0
/l/	0	0	0	0
/r/	0	0	0	0
/w/	0	0	0	0
/j/	0	0	0	0

/m/	0	0	0	0
/n/	0	0	0	0
/ŋ/	0	0	0	0
/h/	0	0	0	0

Appendix 16. Number of tokens featuring each sound per level (activities 1 and 2), and standard deviation (SD)

Activity 1 (with sound n = 1000)						
Level	**/æ/**	**/ʌ/**	**/ɑː/**	**/ə/**	**/s/**	**/z/**
1	5	6	3	7	3	4
2	5	6	3	16	3	2
3	8	2	5	9	4	3
4	7	3	5	8	3	3
5	7	2	4	11	3	6
6	5	2	3	12	6	7
7	9	2	2	6	3	3
8	6	3	2	8	3	3
9	6	1	3	6	4	1
10	10	5	4	5	6	1
SD	1.75	1.81	1.07	3.36	1.23	1.95
Activity 2 (without sound n = 500)						
Level	**/æ/**	**/ʌ/**	**/ɑː/**	**/ə/**	**/s/**	**/z/**
1	5	4	5	10	3	6
2	6	7	2	12	3	5
3	6	2	3	9	4	3
4	8	2	6	8	6	2
5	6	3	3	13	5	3
SD	1.10	2.07	1.64	2.07	1.30	1.64

Appendix 17. Training stimuli for study 2 (activities 1 and 2)

Activity 1 (with sound n = 1000)

	/æ/	Fr.		/ɑː/	Fr.		/ʌ/	Fr.		/ə/	Fr.		/s/	Fr.		/z/	Fr.
1	stamp	7	**1**	argue	6	**1**	cousin	5	**1**	famous	9	**1**	books	7	**1**	flies	7
2	capital	6	**2**	aunt	5	**2**	uncle	4	**2**	Africa	7	**2**	works	7	**2**	buys	6
3	happen	6	**3**	car	5	**3**	young	4	**3**	ago	7	**3**	nice	4	**3**	reads	5
4	actor	5	**4**	dark	5	**4**	comfortable	3	**4**	picture	6	**4**	writes	4	**4**	museum	4
5	flat	5	**5**	answer	4	**5**	hundred	3	**5**	dangerous	5	**5**	cooks	3	**5**	music	4
6	garage	5	**6**	dance	4	**6**	son	3	**6**	dinner	5	**6**	costs	3	**6**	exams	2
7	back	4	**7**	garden	2	**7**	stomach	3	**7**	hairdresser	4	**7**	eats	3	**7**	please	2
8	bad	4	**8**	star	2	**8**	under	3	**8**	October	4	**8**	speaks	3	**8**	arrives	1
9	black	4	**9**	class	1	**9**	bus	2	**9**	pilot	4	**9**	lettuce	2	**9**	words	1
10	chat	4				**10**	come	1	**10**	second	4	**10**	police	2			
11	have	4				**11**	ugly	1	**11**	secretary	4						
12	salad	4							**12**	sugar	4						
13	thanks	3							**13**	chocolate	3						
14	animal	2							**14**	doctor	3						
15	apple	2							**15**	exercise	3						
16	stand	2							**16**	umbrella	3						
17	bag	1							**17**	abroad	2						
									18	colour	2						
									19	Russia	2						
									20	sofa	2						
									21	teacher	2						
									22	autumn	1						
									23	cupboard	1						
									24	mother	1						

Activity 2 (without sound n = 500)

	/æ/	Fr.		/ɑː/	Fr.		/ʌ/	Fr.		/ə/	Fr.		/s/	Fr.		/z/	Fr.
1	back	3	**1**	answer	1	**1**	bus	1	**1**	abroad	3	**1**	books	1	**1**	arrives	1
2	black	3	**2**	argue	4	**2**	come	1	**2**	ago	3	**2**	costs	1	**2**	buys	1
3	capital	2	**3**	aunt	3	**3**	comfortable	1	**3**	autumn	1	**3**	eats	1	**3**	exams	4
4	flat	1	**4**	car	3	**4**	hundred	1	**4**	chocolate	1	**4**	lettuce	3	**4**	flies	2
5	garage	1	**5**	class	3	**5**	son	2	**5**	colour	1	**5**	nice	2	**5**	museum	1
6	happen	3	**6**	dance	1	**6**	stomach	2	**6**	cupboard	1	**6**	police	2	**6**	please	1
7	have	1	**7**	dark	1	**7**	ugly	2	**7**	dangerous	1	**7**	speaks	4	**7**	present	4
8	salad	1	**8**	garden	1	**8**	uncle	1	**8**	dinner	1	**8**	works	3	**8**	reads	3
9	stamp	4	**9**	glasses	1	**9**	under	4	**9**	doctor	1	**9**	writes	4	**9**	words	2
10	stand	2	**10**	star	1	**10**	young	3	**10**	exercise	1						
11	thanks	9							**11**	famous	5						
12	travel	1							**12**	hairdresser	3						
									13	mother	2						
									14	October	1						
									15	picture	4						

16	pilot	2
17	Russia	2
18	second	4
19	secretary	4
20	sofa	2
21	teach<u>er</u>	4
22	umbrella	5

Appendix 18. Written stimuli for the second activity in the identification task in study 2

1. c**o**mfortable	16. c**ou**ple
2. mann**e**r	17. broth**e**r
3. h**u**ndred	18. fam**i**ly
4. g**a**rden	19. c**a**pital
5. Afric**a**	20. **a**gainst
6. **u**nder	21. m**u**ltiple
7. danger**ou**s	22. f**o**rget
8. **a**nswer	23. probl**e**m
9. dinn**e**r	24. dem**a**nd
10. st**o**mach	25. horiz**o**n
11. h**a**ppen	26. hospit**a**l
12. fam**ou**s	27. anoth**e**r
13. **a**ctor	
14. sec**o**nd	
15. Octob**e**r	

Appendix 19. Stimuli for the identification task in study 2

	/æ/		/ʌ/		/ɑː/		/ə/		/s/		/z/
					Familiar						
1	stamp	21	cousin	41	argue	61	famous	81	books	101	flies
2	capital	22	uncle	42	aunt	62	Africa	82	works	102	buys
3	happen	23	young	43	car	63	ago	83	nice	103	reads
4	actor	24	comfortable	44	dark	64	picture	84	writes	104	museum
5	flat	25	hundred	45	answer	65	dangerous	85	cooks	105	music
6	garage	26	son	46	dance	66	dinner	86	costs	106	exams
7	back	27	stomach	47	garden	67	October	87	eats	107	please
8	bad	28	under	48	star	68	pilot	88	speaks	108	arrives
9	black	29	bus	49	class	69	second	89	lettuce	109	words
10	have	30	come	50	glasses	70	sugar	90	police	110	present
					Novel						
11	brag	31	jump	51	farm	71	manner	91	ceiling	111	zombie
12	jazz	32	money	52	park	72	hospital	92	centre	112	zebra
13	flag	33	blood	53	past	73	another	93	December	113	busy
14	gang	34	tough	54	smart	74	against	94	decent	114	easy
15	happy	35	couple	55	laugh	75	problem	95	discipline	115	fizzy
16	fax	36	run	56	demand	76	forget	96	recipe	116	amazing
17	hang	37	lunch	57	dart	77	prison	97	peace	117	rose
18	glad	38	touch	58	bark	78	horizon	98	niece	118	amuse
19	gas	39	multiple	59	balm	79	brother	99	ice	119	these
20	fan	40	sun	60	garlic	80	family	100	pace	120	cheese

	Distractors
	/ʃ/
121	shoe
122	she
123	issue
124	mission
125	bush

Note: All items were obtained from the pronunciation dictionaries explained in section 3.1.1, except for *books*, *works*, *writes*, *cooks*, *costs*, *eats*, and *speaks* for /s/, and *flies*, *buys*, *reads*, *exams*, *arrives*, and *words* for /z/.

Appendix 20. Stimuli for the sentence-reading task (study 2)

	/æ/		/ʌ/		/ɑː/		/ə/		/z/
				Familiar stimuli					
1	stamp	16	cousin	31	argue	46	famous	61	flies
2	capital	17	uncle	32	aunt	47	Africa	62	buys
3	happen	18	young	33	car	48	ago	63	reads
4	actor	19	comfortable	34	dark	49	picture	64	museum
5	flat	20	hundred	35	answer	50	dangerous	65	music
6	garage	21	son	36	dance	51	dinner	66	exams
7	back	22	stomach	37	garden	52	October	67	please
8	bad	23	under	38	star	53	pilot	68	arrives
9	black	24	bus	39	class	54	second	69	words
10	have	25	come	40	glasses	55	sugar	70	present
				Novel stimuli					
11	fan	26	gum	41	far	56	father	71	frozen
12	hang	27	bug	42	father	57	lemon	72	magazine
13	lack	28	run	43	half	58	oven	73	zebra
14	sad	29	brother	44	large	59	student	74	lazy
15	anger	30	fun	45	guitar	60	camera	75	size

Carrier sentences

1	The car is in the garage.	**23**	The capital is comfortable.
2	The answer is in the garden.	**24**	He reads a hundred books in the summer.
3	The picture is in the museum.	**25**	He arrives in the second car.
4	The bus is black.	**26**	My brother has a camera.
5	Sugar is bad.	**27**	The size is large.
6	The pilot is in Africa.	**28**	The lemon is frozen.
7	The actor is famous.	**29**	Anger isn't fun.
8	My aunt is young.	**30**	He's a student. He plays the guitar.
9	My uncle is back.	**31**	My father is lazy.
10	The dark one is dangerous.	**32**	That's not a bug, it's a zebra.
11	We argue with my cousin.	**33**	They lack half.
12	The exams are in the class.	**34**	The oven is far.
13	His son is having dinner.	**35**	Let's hang the fan.
14	I have a flat.	**36**	He was sad and run.
15	They bought him a present. It's a star.	**37**	The gum is in the magazine.
16	Please, don't go in October.		
17	The music is for the dance.		
18	It happened a month ago.		
19	The words are in the stamp.		
20	The flies have come.		
21	He buys new glasses.		
22	It's under the stomach.		

Appendix 21. Mean scores and percentage of improvement (%) for each participant in the production test for each sound across tasks, including the total points for each sound across tasks (max)

		/æ/			/ʌ/			/ə/			/z/		
	max	22			21			23			22		
Gr	**ppt**	**pre**	**post**	**%**	**pre**	**post**	**%**	**pre**	**post**	**%**	**pre**	**post**	**%**
1	**3**	7	8	4.5	7	10	14.3	11	11	0.0	0	1	4.5
1	**4**	0	5	22.7	6	9	14.3	0	7	30.4	0	0	0.0
1	**5**	1	2	4.5	6	11	23.8	4	8	17.4	0	1	4.5
1	**6**	1	5	18.2	7	9	9.5	4	4	0.0	0	0	0.0
1	**9**	0	0	0.0	1	3	9.5	0	0	0.0	0	0	0.0
1	**11**	4	10	27.3	11	8	-14.3	7	7	0.0	1	1	0.0
1	**12**	0	0	0.0	5	8	14.3	2	3	4.3	0	0	0.0
1	**13**	4	3	-4.5	2	3	4.8	17	19	8.7	0	1	4.5
1	**14**	8	10	9.1	9	15	28.6	6	11	21.7	3	1	-9.1
1	**17**	4	6	9.1	9	16	33.3	14	16	8.7	1	1	0.0
1	**23**	5	6	4.5	6	10	19.0	1	4	13.0	1	0	-4.5
1	**25**	2	14	54.5	4	7	14.3	4	9	21.7	17	20	13.6
1	**31**	2	2	0.0	7	8	4.8	3	4	4.3	0	0	0.0
1	**32**	4	8	18.2	10	12	9.5	7	8	4.3	0	9	40.9
1	**33**	0	0	0.0	6	8	9.5	1	1	0.0	0	1	4.5
1	**34**	8	16	36.4	6	10	19.0	12	16	17.4	2	5	13.6
1	**35**	0	5	22.7	6	10	19.0	5	7	8.7	1	4	13.6
1	**36**	6	11	22.7	5	6	4.8	7	7	0.0	18	18	0.0
1	**37**	5	13	36.4	8	13	23.8	6	12	26.1	7	11	18.2
1	**39**	0	1	4.5	3	5	9.5	1	2	4.3	1	1	0.0
1	**42**	1	1	0.0	8	10	9.5	4	4	0.0	3	4	4.5
1	**47**												
1	**50**	1	1	0.0	3	7	19.0	4	2	-8.7	3	0	-13.6
1	**54**	3	1	-9.1	4	10	28.6	1	2	4.3	3	9	27.3
1	**60**	2	11	40.9	5	4	-4.8	1	4	13.0	0	1	4.5
1	**84**	0	3	13.6	6	9	14.3	5	9	17.4	0	2	9.1
1	**91**	1	0	-4.5	5	8	14.3	1	4	13.0	1	4	13.6
2	**7**	0	2	9.1	6	6	0.0	0	2	8.7	0	1	4.5
2	**8**	4	4	0.0	1	0	-4.8	2	4	8.7	1	0	-4.5
2	**10**	5	6	4.5	8	11	14.3	5	7	8.7	0	2	9.1
2	**15**	0	0	0.0	4	3	-4.8	2	3	4.3	1	1	0.0
2	**16**	0	1	4.5	8	10	9.5	4	6	8.7	0	1	4.5
2	**18**	0	1	4.5	5	6	4.8	9	6	-13.0	2	3	4.5
2	**19**	0	4	18.2	4	6	9.5	1	2	4.3	3	9	27.3
2	**21**	1	5	18.2	7	10	14.3	3	2	-4.3	0	0	0.0
2	**22**	9	15	27.3	11	14	14.3	13	13	0.0	2	6	18.2
2	**24**	4	2	-9.1	6	10	19.0	8	8	0.0	3	6	13.6
2	**26**	13	13	0.0	8	10	9.5	13	15	8.7	2	1	-4.5
2	**30**	1	0	-4.5	8	8	0.0	1	3	8.7	0	0	0.0

2	**38**	0	0	0.0	9	8	-4.8	0	0	0.0	0	0	0.0
2	**40**	3	3	0.0	2	3	4.8	3	2	-4.3	0	0	0.0
2	**43**	2	4	9.1	7	9	9.5	3	7	17.4	2	2	0.0
2	**44**	1	2	4.5	8	11	14.3	4	6	8.7	1	0	-4.5
2	**45**	1	0	-4.5	4	9	23.8	1	1	0.0	0	0	0.0
2	**46**	4	1	-13.6	7	7	0.0	3	3	0.0	0	0	0.0
2	**49**	0	0	0.0	7	9	9.5	1	2	4.3	1	2	4.5
2	**53**	0	6	27.3	12	6	-28.6	6	6	0.0	0	0	0.0
2	**56**	14	18	18.2	6	9	14.3	16	18	8.7	6	9	13.6
2	**58**	1	1	0.0	7	10	14.3	4	2	-8.7	3	1	-9.1
2	**59**	1	3	9.1	8	5	-14.3	1	1	0.0	1	3	9.1
2	**90**	0	1	4.5	6	7	4.8	4	4	0.0	3	3	0.0
2	**96**	0	0	0.0	2	3	4.8	0	0	0.0	2	2	0.0
2	**100**	0	0	0.0	1	3	9.5	0	0	0.0	0	0	0.0
2	**101**	5	7	9.1	7	9	9.5	1	3	8.7	0	0	0.0
3	**1**	0	2	9.1	8	8	0.0	9	11	8.7	11	11	0.0
3	**2**	1	5	18.2	1	3	9.5	0	1	4.3	0	1	4.5
3	**62**	0	0	0.0	4	6	9.5	1	1	0.0	1	0	-4.5
3	**63**	0	1	4.5	5	4	-4.8	0	1	4.3	0	0	0.0
3	**75**	0	13	59.1	13	16	14.3	9	16	30.4	13	23	45.5
3	**78**	1	0	-4.5	3	3	0.0	0	1	4.3	0	0	0.0
3	**80**	6	11	22.7	13	15	9.5	12	15	13.0	0	0	0.0
3	**82**	9	16	31.8	12	15	14.3	20	18	-8.7	3	5	9.1
3	**83**	3	10	31.8	7	12	23.8	5	5	0.0	2	1	-4.5
3	**95**												

Appendix 22. Mean scores for pre- and post-test and the degree of improvement made (in percentages) by individual participants in the sentence-reading task (study 2

		/æ/			/ʌ/			/ə/			/z/		
group	**ppt**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**	**pre**	**post**	**imp**
1	**3**	28.6	35.7	7.1	50.0	57.1	7.1	53.3	53.3	0.0	0.0	0.0	0.0
1	**4**	0.0	14.3	14.3	14.3	35.7	21.4	0.0	33.3	33.3	0.0	0.0	0.0
1	**5**	0.0	7.1	7.1	28.6	42.9	14.3	20.0	33.3	13.3	0.0	0.0	0.0
1	**6**	0.0	14.3	14.3	28.6	64.3	35.7	20.0	20.0	0.0	0.0	0.0	0.0
1	**9**	0.0	0.0	0.0	7.1	7.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	**11**	28.6	42.9	14.3	50.0	28.6	-21.4	40.0	33.3	-6.7	0.0	0.0	0.0
1	**12**	0.0	0.0	0.0	28.6	21.4	-7.1	6.7	6.7	0.0	0.0	0.0	0.0
1	**13**	21.4	14.3	-7.1	7.1	7.1	0.0	66.7	73.3	6.7	0.0	0.0	0.0
1	**14**	28.6	35.7	7.1	50.0	92.9	42.9	26.7	46.7	20.0	13.3	0.0	-13.3
1	**17**	7.1	14.3	7.1	42.9	71.4	28.6	53.3	66.7	13.3	0.0	0.0	0.0
1	**23**	21.4	21.4	0.0	35.7	50.0	14.3	6.7	20.0	13.3	6.7	0.0	-6.7
1	**25**	7.1	50.0	42.9	28.6	35.7	7.1	13.3	33.3	20.0	86.7	100.0	13.3
1	**31**	14.3	7.1	-7.1	35.7	35.7	0.0	20.0	20.0	0.0	0.0	0.0	0.0
1	**32**	21.4	35.7	14.3	64.3	64.3	0.0	40.0	40.0	0.0	0.0	46.7	46.7
1	**33**	0.0	0.0	0.0	35.7	35.7	0.0	6.7	6.7	0.0	0.0	0.0	0.0
1	**34**	14.3	50.0	35.7	35.7	50.0	14.3	46.7	66.7	20.0	0.0	20.0	20.0
1	**35**	0.0	28.6	28.6	35.7	64.3	28.6	13.3	33.3	20.0	0.0	13.3	13.3
1	**36**	21.4	42.9	21.4	35.7	35.7	0.0	26.7	20.0	-6.7	73.3	73.3	0.0
1	**37**	21.4	64.3	42.9	42.9	64.3	21.4	20.0	40.0	20.0	33.3	46.7	13.3
1	**39**	0.0	0.0	0.0	7.1	21.4	14.3	0.0	0.0	0.0	0.0	0.0	0.0
1	**42**	0.0	7.1	7.1	28.6	28.6	0.0	20.0	26.7	6.7	0.0	0.0	0.0
1	**47**												
1	**50**	0.0	0.0	0.0	14.3	21.4	7.1	13.3	6.7	-6.7	0.0	0.0	0.0
1	**54**	14.3	7.1	-7.1	21.4	50.0	28.6	6.7	6.7	0.0	13.3	46.7	33.3
1	**60**	0.0	50.0	50.0	35.7	28.6	-7.1	6.7	20.0	13.3	0.0	6.7	6.7
1	**84**	0.0	0.0	0.0	28.6	35.7	7.1	20.0	33.3	13.3	0.0	6.7	6.7
1	**91**	0.0	0.0	0.0	28.6	35.7	7.1	0.0	13.3	13.3	6.7	13.3	6.7
2	**7**	0.0	0.0	0.0	21.4	14.3	-7.1	0.0	13.3	13.3	0.0	0.0	0.0
2	**8**	7.1	14.3	7.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	**10**	21.4	28.6	7.1	42.9	35.7	-7.1	20.0	26.7	6.7	0.0	6.7	6.7
2	**15**	0.0	0.0	0.0	21.4	14.3	-7.1	6.7	13.3	6.7	0.0	0.0	0.0
2	**16**	0.0	0.0	0.0	50.0	50.0	0.0	20.0	26.7	6.7	0.0	0.0	0.0
2	**18**	0.0	7.1	7.1	21.4	21.4	0.0	40.0	33.3	-6.7	13.3	13.3	0.0
2	**19**	0.0	14.3	14.3	14.3	21.4	7.1	6.7	6.7	0.0	13.3	33.3	20.0
2	**21**	7.1	28.6	21.4	21.4	35.7	14.3	6.7	6.7	0.0	0.0	0.0	0.0
2	**22**	28.6	78.6	50.0	42.9	57.1	14.3	46.7	46.7	0.0	6.7	13.3	6.7
2	**24**	21.4	7.1	-14.3	35.7	57.1	21.4	20.0	20.0	0.0	20.0	33.3	13.3
2	**26**	57.1	57.1	0.0	50.0	64.3	14.3	60.0	60.0	0.0	6.7	0.0	-6.7
2	**30**	7.1	0.0	-7.1	28.6	35.7	7.1	0.0	6.7	6.7	0.0	0.0	0.0
2	**38**	0.0	0.0	0.0	28.6	28.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0

2	**40**	7.1	7.1	0.0	7.1	14.3	7.1	13.3	13.3	0.0	0.0	0.0	0.0
2	**43**	0.0	14.3	14.3	28.6	28.6	0.0	13.3	26.7	13.3	0.0	0.0	0.0
2	**44**	0.0	0.0	0.0	35.7	42.9	7.1	13.3	20.0	6.7	0.0	0.0	0.0
2	**45**	0.0	0.0	0.0	21.4	42.9	21.4	6.7	6.7	0.0	0.0	0.0	0.0
2	**46**	21.4	0.0	-21.4	42.9	35.7	-7.1	20.0	20.0	0.0	0.0	0.0	0.0
2	**49**	0.0	0.0	0.0	35.7	35.7	0.0	6.7	13.3	6.7	0.0	0.0	0.0
2	**53**	0.0	35.7	35.7	50.0	35.7	-14.3	20.0	26.7	6.7	0.0	0.0	0.0
2	**56**	64.3	85.7	21.4	28.6	35.7	7.1	73.3	80.0	6.7	13.3	20.0	6.7
2	**58**	0.0	0.0	0.0	35.7	42.9	7.1	26.7	6.7	-20.0	13.3	0.0	-13.3
2	**59**	7.1	21.4	14.3	35.7	35.7	0.0	6.7	6.7	0.0	6.7	20.0	13.3
2	**90**	0.0	7.1	7.1	21.4	21.4	0.0	13.3	20.0	6.7	13.3	20.0	6.7
2	**96**	0.0	0.0	0.0	7.1	14.3	7.1	0.0	0.0	0.0	6.7	6.7	0.0
2	**100**	0.0	0.0	0.0	0.0	14.3	14.3	0.0	0.0	0.0	0.0	0.0	0.0
2	**101**	21.4	21.4	0.0	28.6	35.7	7.1	6.7	13.3	6.7	0.0	0.0	0.0
3	**1**	0.0	14.3	14.3	28.6	21.4	-7.1	40.0	53.3	13.3	40.0	40.0	0.0
3	**2**	0.0	21.4	21.4	0.0	7.1	7.1	0.0	6.7	6.7	0.0	0.0	0.0
3	**62**	0.0	0.0	0.0	21.4	35.7	14.3	6.7	6.7	0.0	0.0	0.0	0.0
3	**63**	0.0	0.0	0.0	21.4	21.4	0.0	0.0	6.7	6.7	0.0	0.0	0.0
3	**75**	0.0	71.4	71.4	57.1	64.3	7.1	40.0	73.3	33.3	60.0	113.3	53.3
3	**78**	7.1	0.0	-7.1	14.3	14.3	0.0	0.0	6.7	6.7	0.0	0.0	0.0
3	**80**	21.4	50.0	28.6	50.0	64.3	14.3	66.7	73.3	6.7	0.0	0.0	0.0
3	**82**	35.7	71.4	35.7	50.0	57.1	7.1	86.7	80.0	-6.7	13.3	26.7	13.3
3	**83**	14.3	42.9	28.6	35.7	50.0	14.3	26.7	26.7	0.0	0.0	0.0	0.0
3	**95**												

Appendix 23. Initial questionnaire study 2

Encuesta inicial

***Obligatorio**

Sexo *

- Hombre
- Mujer

Edad * __________

¿Qué carrera/estudios oficiales estas cursando actualmente? *

- Grado en Estudios Ingleses
- Grado en Traducción e Interpretación
- Grado en Lengua y Literatura Españolas
- Grado en Medicina
- Otro: __________

¿Cuánto tiempo pasas al día en internet? *

- Menos de 15 minutos
- Entre 15 minutos y 30
- Entre 30 minutos y 1 hora
- Entre 1 y 2 horas
- Más de 2 horas

¿Desde qué dispositivo sueles conectarte a internet? *

- Desde el móvil (por la calle)
- Desde el móvil (en casa)
- Tablet
- Ordenador
- Otro: __________

¿Qué redes sociales usas? *

Se entiende por 'usarlas' a que tienes cuenta y las visitas de manera diaria. (Puedes marcar más de una opción)

- Tuenti
- Twitter
- Facebook
- LinkedIn
- Google+
- Instagram
- No uso redes sociales
- Otro: __________

¿Qué modelo de teléfono tienes? *

Por favor, especifica marca y modelo (ejemplo: Samsung Galaxy Ace 2)

____

¿Es un smartphone? *

Un smartphone es un teléfono con pantalla táctil que permite descargar aplicaciones

- Si
- No

¿Qué sistema operativo tiene? *

- Android
- iOS (iPhone)
- Windows
- Otro: __________

¿Cuánto tiempo utilizas tu teléfono al día? *
(para hablar, chatear, navegar por internet, estar en redes sociales, etc.)

- Menos de 15 minutos
- Entre 15 y 30 minutos
- Entre 30 minutos y 1 hora
- Entre 1 y 2 horas
- Más de 2 horas

¿Has utilizado alguna vez aplicaciones móviles con fines educativos? *

- Sí
- No

Si la respuesta anterior es afirmativa, por favor, indica para qué asignatura/tema.
____

¿Cuál es tu postura sobre esta afirmación? *
Me gustaría utilizar aplicaciones móviles para aprender idiomas.
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO

¿Cuál es tu postura sobre esta afirmación? *
Considero que una aplicación móvil puede resultar útil para aprender idiomas.
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO

¿Cuál es tu postura sobre esta afirmación? *
Considero que una aplicación móvil puede ser útil para ayudarme a mejorar la pronunciación
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO

¿Cuánto tiempo estarías dispuesto a utilizar una aplicación móvil para aprender idiomas? *

- Menos de 15 minutos al día
- Entre 15 y 30 minutos al día
- Entre 30 minutos y 1 hora al día
- Más de 1 hora al día
- Otro: __________

¿Cómo piensas que un nativo valoraría tu nivel de inglés después de escucharte hablar durante un rato? *

- Completamente bilingüe o casi bilingüe
- Muy alto
- Avanzado
- Intermedio
- Elemental
- Principiante
- Otro: __________

¿Cuánto tiempo estás expuesto al inglés habitualmente? *
(Incluyendo el tiempo en clase, viendo la televisión, escuchando la radio en inglés, etc.)

- Menos de 1 hora al día
- Entre 1 y 2 horas al día
- Más de 2 horas al día
- Entre 1 y 2 horas a la semana
- Otro: __________

SIN CONTAR LAS HORAS DE CLASE, ¿cuánto tiempo estás expuesto al inglés en tu tiempo libre? *
(Viendo la televisión/cine, escuchando la radio, con videojuegos, con amigos, etc.)

- Menos de 1 hora al día
- Entre 1 y 2 horas al día
- Más de 2 horas al día
- Entre 1 y 2 horas a la semana.
- Otro: __________

¿Cómo? *
(Puedes marcar varias opciones)
- Viendo la televisión
- Escuchando la radio
- Escuchando música
- Hablando con amigos
- Con videojuegos
- Viendo entrevistas/similares en YouTube
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Presto atención a cómo la gente pronuncia en inglés y me comparo con ellos (La forma en que acentúan las palabras, cómo pronuncian los sonidos, la entonación que usan, etc.)
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, me doy cuenta.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy consciente de cuál es el problema.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Cuando cometo fallos de pronunciación, soy capaz de corregir el problema yo mismo.
Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

De ser afirmativa la anterior, ¿cómo?
(Puedes elegir varias opciones e incluir otras que no estén en la lista)
- Con un diccionario online
- Con la función de 'pronunciar' del traductor de Google o similar
- Le pregunto a alguien
- No hago nada
- Otro: __________

¿Cuál es tu postura sobre la siguiente afirmación? *
Para mí es muy importante conseguir una pronunciación nativa, que alguien me oiga y no sepa si soy inglés o no.
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO

¿Qué modelo de inglés prefieres? *
- Británico
- Americano
- Otro: __________

¿Por qué prefieres uno u otro? *

____

¿Cómo describirías tu nivel de inglés? *
(Siendo realista, nivel obtenido, no cursando actualmente) Los niveles entre paréntesis son los descriptores del Marco Europeo de Referencia
- Completamente bilingüe o casi bilingüe (nivel C2)
- Muy alto (nivel C1)
- Avanzado (nivel B2)
- Intermedio (nivel B1)
- Elemental (nivel A2)
- Principiante (nivel A1)
- Otro: __________

¿Tienes algún título oficial que certifique el nivel de inglés que mencionas arriba? *
- Sí
- No

Appendix 24. Final questionnaire study 2

Encuesta final

***Obligatorio**
¿Cuántas HORAS has pasado estudiando inglés a lo largo de tu vida? *
Ten en cuenta que en secundaria y bachillerato se estudia inglés una media de 105 horas anuales (3 horas a la semana durante 35 semanas -quitando festivos y vacaciones). Por favor, haz un cálculo (honesto) aproximado. Suma si has asistido a Escuelas de Idiomas, academias, etc...

____

Antes de este estudio, ¿has utilizado otras aplicaciones móviles con fines educativos? *

- Sí
- No

Si la respuesta anterior es afirmativa, por favor, indica para qué asignatura/s o temas y el tipo de actividad realizada.
Ejemplo: para matemáticas, para aprender a hacer divisiones.

____

Por favor, indica tu postura sobre la siguiente afirmación *
Después de esta experiencia me gustaría utilizar aplicaciones móviles para aprender inglés
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
En general, la aplicación me ha parecido útil
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Considero que el contenido visto en la aplicación me va a resultar útil en mi uso cotidiano del inglés
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Creo que la aplicación me ha ayudado a DISTINGUIR mejor los sonidos estudiados
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Creo que la aplicación me ha ayudado a PRONUNCIAR mejor los sonidos estudiados
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
He disfrutado utilizando la aplicación
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Los aspectos tratados me parecen interesantes
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
La aplicación me ha parecido entretenida, no me aburría
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
La aplicación me ha parecido monótona
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
¿Has tenido algún problema técnico mientras utilizabas la aplicación? *

- Sí
- No

De ser afirmativa la anterior, por favor, indica cuál y cómo lo has solucionado

____

Si has visto parte de los contenidos en algún sitio además del estudio (durante la duración del mismo), por favor indica dónde y concretamente qué aspectos has visto
(Ejemplo: "He visto el sonido de "sh" en 'she'; en clase, en la Escuela de Idiomas, etc.)

____

Por favor, indica tu postura sobre la siguiente afirmación *
La aplicación me ha parecido fácil de manejar
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
El numero de palabras ejemplo para cada sonido en el cuadro de fonemas me parece suficiente
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
El numero de palabras ejemplo en las actividades me parece suficiente
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Si has necesitado utilizar las pantallas de ayuda de la aplicación, ¿has conseguido resolver la duda que tuvieras con la información facilitada? *
Si quieres añadir información específica, por favor marca la opción 'otro' y añade ahí tus comentarios

- Sí
- No
- No he utilizado las pantallas de ayuda
- Otro __________

Por favor, indica tu postura sobre la siguiente afirmación *
El diseño de la aplicación me ha parecido atractivo
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Pienso seguir utilizando la aplicación una vez que acabe el estudio
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Si sacan niveles y actividades nuevas estoy dispuesto a usarlos
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Las actividades me han parecido difíciles
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
El feedback que me indicaba si mis elecciones eran correctas o no me ha parecido adecuado
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Por favor, indica tu postura sobre la siguiente afirmación *
Cuando daba una respuesta incorrecta el feedback me ha ayudado a saber por qué mi respuesta era incorrecta
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
Mientras jugabas, ¿has tenido alguna duda que no hayas podido resolver usando la App?
En caso afirmativo, por favor indica cuál.

- Sí
- No
- Otro: __________

¿Te ha parecido apropiada la duración del estudio (10 días)? *

- Si, el tiempo justo.
- Si, pero me hubieran venido bien unos días más.
- Un poco largo, con unos días menos habría estado bien.
- Otro: __________

¿Habías utilizado antes un libro de Oxford con un cuadro de símbolos fonéticos similar? *

- Sí
- No

¿Qué actividad te ha parecido más útil? *

- Cuadro de fonemas
- Actividad 1 (con audio)
- Actividad 2 (sin audio)

Por favor, indica el grado de utilidad que le ves al cuadro de fonemas de la app
NADA ÚTIL 1 2 3 4 5 MUY ÚTIL
Por favor, indica el grado de utilidad que le ves a la actividad 1 (con sonido) de la app
NADA ÚTIL 1 2 3 4 5 MUY ÚTIL
Por favor, indica el grado de utilidad que le ves a la actividad 2 (sin sonido) de la app
NADA ÚTIL 1 2 3 4 5 MUY ÚTIL
¿Has utilizado la función de grabación para repetir después del modelo? *

- Sí
- No

¿Has repetido en voz alta las palabras después del modelo para intentar imitar su pronunciación? *

- Siempre
- Con frecuencia
- A veces
- Casi nunca
- Nunca

Si yo no te hubiera dicho cómo utilizar la app, ¿cómo la habrías utilizado tú? *
(¿habrías utilizado las mismas actividades, el mismo orden, etc.?)

____

¿Conocías ya el vocabulario que sale en la App? *

- Sí
- No
- No lo sé, no me he fijado

Por favor, indica tu postura sobre la siguiente afirmación (contesta ÚNICAMENTE si te has fijado en el vocabulario que salía)
El vocabulario que sale en la app me ha parecido útil
Totalmente EN DESACUERDO 1 2 3 4 5 Totalmente DE ACUERDO
¿Crees que se repiten las palabras que salen en la App? *

- No, nunca.
- A veces.
- Bastante.
- Mucho.
- No lo sé, no me he fijado

¿Te ha parecido apropiado el tiempo que tenías que usar la App cada día? *
(La práctica con el cuadro de fonemas más los 10 niveles de cada actividad)

- Sí
- No, ha sido demasiado.
- No, ha sido poco.
- Otro: __________

Por favor, indica cuánto tiempo diario con la App consideras necesario para aprender los sonidos que hemos visto *

____

¿Has mirado por tu cuenta alguno de los sonidos que yo NO pedí, pero que estaban en el cuadro de sonidos? *
(De ser afirmativo, si te acuerdas indica cuál. Ejemplo: la "s" de SHe).

____

En los resúmenes de nivel (donde te indica los fallos y aciertos que has tenido), ¿te fijabas en los fallos que tenías con intención de mejorar? *

- Sí, siempre.
- Sí, a menudo.
- Sí, a veces.
- Rara vez.
- Nunca.

¿Qué recomendaciones me darías para mejorar futuras ediciones de este curso? *
(sobre la duración, cómo utilizar las actividades, u otras cosas que se te ocurran)

____

Pensando en cómo está diseñada la aplicación, ¿cambiarías algo? *
(Algo que notes que le falta, algún fallo que hayas visto y quieras comentar).

____

¿Pagarías 5,50€ por esta aplicación? *

- Si, merece la pena.
- No, es demasiado caro.
- No, porque nunca compro aplicaciones, pero si comprase me parece un precio razonable.
- Otro: __________

¿Qué precio le pondrías tú a esta aplicación? *
(Teniendo el cuenta el precio de otras aplicaciones del mercado)

____

Por favor, indica qué momento del día preferías para utilizar la app y por qué *

____

Por favor, si hay algo que quieras comentar o sugerir, hazlo aquí:

____

Appendix 25. Training stimuli for study 3 arranged in chronological order

1. Item: **Pariah**
Bridget Jones is a love pariah no more. Careful with 'pariah', 'riah' being pronounced as 'Mariah' or 'fire'. http://youtu.be/AB4bh1CpZyo
2. Item: **catholic**
Pay attention to the word CAtholic (católico). Spaniards tend to say caTHOlic, but the stress is on the first syllable http://www.goear.com/listen/3ef52db/catholic-colour
3. Item: **aisle**
Phoebe chooses Joey to 'walk her down the AISLE'! Careful! the 's' in 'aisle' is not pronounced! http://www.youtube.com/watch?v=1vPr0dQtn7o
4. Item: **Greenwich**
You've probably seen the abbreviation 'GMT' many times in your life (Greenwich Mean Time). Watch this clip and pay attention to how GREENWICH is pronounced: the 'w' is silent! http://youtu.be/l6gosssl0ts
5. Item: **Politics**
Today's word may be really useful these days: POlitics. Careful with the stress! it's not poLItics, but POlitics! http://youtu.be/QXWXXMrkRJE
6. Item: **Gauge**
Do you know what a GAUGE is? Watch the first 30 secs & notice 'au', pronounced as 'a' in game & 'ge' as 'J' in John http://youtu.be/euIL8M2d2L0
8. Item: **Charlatan**
Pay attention to the word 'Charlatan' (charlatán/embustero), the 'ch' being pronounced as the 'sh' in 'she'. http://youtu.be/Uu3vyFUmDDs
10. Item: **Heir**
As 'The Smiths' say: 'I am the son and the heir'. CAREFUL with the word 'heir'! (heredero) The H is not pronounced!! http://www.youtube.com/watch?v=qghCe_5Kea0
11. Item: **Buckingham**
How do you pronounce the word'Buckingham'? [1/1] Careful! The 'h' is not pronounced! Listen to this link from the BBC about 'Buckingham Palace' and try to imitate it. http://youtu.be/tP4d7XpstYQ
12. Item: **Guitarist**
Are you familiar with Jimy Hendrix? He was a great 'guiTArist' (careful! Unlike Spanish, stress on the 2nd syllable). http://youtu.be/RcKKz_Sp7ik
13. Item: **Half**
Listen to this Oasis's song. Notice how the L in haLf is silent! Spaniards often pronounce it but it's NOT pronounced! http://www.youtube.com/watch?v=-ap4n_3bY_Y
14. Item: **Intermittent**.
Do you know what IED is? Beware of the pronunciation of interMIttent, many people stress 'ter' but the stress is on 'mi'! http://youtu.be/6V827tIi-h4
15. Item: **Stephen**
Notice how Kesha pronounces the name 'Stephen'. Many Spaniards say the 'ph' with an /f/ but it is a /v/!! as in 'vein'. http://www.youtube.com/watch?v=rYbTizdXG3w
16. Item: **Archives**
Listen to the 'ch' in 'archives' (archivo). Careful! It's the same sound as the 'k' in 'key' or the 'c' in 'come'! http://www.goear.com/listen/922fed4/archives-bbc
17. Item: **Graham**
How do you pronounce 'Graham'? Don't say it pronouncing the 'h' as the 'j' in Jamón, the 'h' is not pronounced! http://youtu.be/gi_MHn8ZGis
18. Item: **Terrorism**
Listen to the first 40 seconds of this video and pay attention to TErrorism (stress on

the 1st syllable, not like Spanish terroRIsmo!).
http://www.youtube.com/watch?v=RXF9maO9_wE

19. Item: **Chalk**
Ready for strange addictions? This girl is addicted to chalk! (tiza). Careful with the L, it's NOT PRONOUNCED!
http://www.youtube.com/watch?v=jLl5X9waPkg&sns=em
20. Item: **sword**
Pay attention to the pronunciation of 'sword'. Notice that the 'w' is not pronounced!
http://www.youtube.com/watch?v=d1RYsZc9KJE
21. Item: **cough**
Watch up to min 1.30 and **p**ay attention to the worth 'cough' (toser). The GH is pronounced with /f/ (as the 'f' in 'fly'). Did you know about this vaccine?
http://www.youtube.com/watch?v=q0_DujBo3jQ
22. Item: **mishap**
Did you know the word 'mishap'? (percance)Careful! The 'h' IS pronounced!a mishap in the musical version of Spiderman
http://www.youtube.com/watch?v=h1Uau120T_M
23. Item: **trough**
    a. (1/2) We are happy to introduce the new, the amazing, the latest...water trough! Careful! 'trough' (abrevadero) is NOT 'through'!
    b. (2/2) Two problematic sounds:'ou', pronounced as the 'o' in 'hot' & 'gh' as the 'f' in 'fly' http://youtu.be/hKYCkzBn2lo
24. Item: **bombing**
In this piece of news you have 2 examples of the word 'bombing'. CAREFUL! The 'b' in the middle is NOT pronounced! http://www.youtube.com/watch?v=-63qtqE_dGk

Appendix 26. Pre- (and post-) test stimuli for study 3

house	clean	*heir*	MODERNISM	northern
castle	*chalk*	ORGANISE	cough	*bristle*
TERRORISM	POLITICS	shepherd	*Graham*	drive
monarch	social	grass	cloud	ELITE
talk	money	*aisle*	ATHLETE	archives
tree	*receipt*	PACIFY	stop	*calm*
chemical	RHETORIC	southern	*debt*	PARIAH
Buckingham	trough	greenhouse	CLARIFY	Thames
REALISE	computer	INTERMITTENT	although	mountain
sugar	*half*	charlatan	table	*would*
MOTIVATE	Stephen	sofa	*fasten*	LUNATIC
worthy	lake	*bomb*	FLORIDA	laugh
river	*coup*	COMPOUND	sure	bravery
should	PROPRIETOR	crucial	kitchen	*combing*
GUITARIST	weapon	wood	*yacht*	mishap
cater	water	*Greenwich*	COMMENTARY	gauge
t-shirt	*sword*	TRANQUIL	spatial	book
hasten	CANTERBURY	savour	sky	*island*
CATHOLIC	Thai	pen	*comb*	steak
status	rubber	bombing	Munich	COMMITTEE

Note: Sound-grapheme correspondences (underlined items), silent letters (items in italics), and unexpected stress patterns (items in capitals). The rest of items represent distractors.

Appendix 27. Participants (PPT), learning potential (LPOT), amount of learning (TOT.L) and retention (TOT.R), percentage of learning (%L), and percentages of retention as compared to the number of words learnt (%R) and in relation to the total learning potential for each participant (%R/LPOT)

PPT	Group	LPOT	TOT.L	TOT.R	% L	% R	% R/LPOT
19	1	20	1	0	5.0	0.0	0.0
21	1	21	1	0	4.8	0.0	0.0
26	1	22	0	0	0.0		0.0
40	1	19	3	0	15.8	0.0	0.0
43	1	22	0	0	0.0		0.0
44	1	15	1	0	6.7	0.0	0.0
45	1	20	0	0	0.0		0.0
46	1	18	0	0	0.0		0.0
47	1	21	1	1	4.8	100.0	4.8
48	1	20	1	0	5.0	0.0	0.0
49	1	20	0	0	0.0		0.0
50	1	21	0	0	0.0		0.0
52	1	22	0	0	0.0		0.0
95	1	15	1	0	6.7	0.0	0.0
96	1	21	1	0	4.8	0.0	0.0
97	1	19	0	0	0.0		0.0
98	1	19	2	0	10.5	0.0	0.0
100	1	21	0	0	0.0		0.0
101	1	22	1	0	4.5	0.0	0.0
102	1	15	1	1	6.7	100.0	6.7
104	1	21	0	0	0.0		0.0
107	1	19	0	0	0.0		0.0
108	1	22	0	0	0.0		0.0
109	1	21	1	0	4.8	0.0	0.0
110	1	21	0	0	0.0		0.0
111	1	22	1	0	4.5	0.0	0.0
113	1	21	0	0	0.0		0.0
116	1	21	1	0	4.8	0.0	0.0
121	1	17	1	0	5.9	0.0	0.0
122	1	21	2	0	9.5	0.0	0.0
163	1	22	0	0	0.0		0.0
164	1	21	0	0	0.0		0.0
172	1	22	1	0	4.5	0.0	0.0
174	1	22	2	0	9.1	0.0	0.0
175	1	18	0	0	0.0		0.0
176	1	19	1	0	5.3	0.0	0.0
177	1	17	1	0	5.9	0.0	0.0
178	1	20	0	0	0.0		0.0
184	1	21	0	0	0.0		0.0
186	1	19	0	0	0.0		0.0
187	1	22	0	0	0.0		0.0

188	1	22	0	0	0.0		0.0
193	1	22	0	0	0.0		0.0
31	2	18	10	8	55.6	80.0	44.4
32	2	22	6	5	27.3	83.3	22.7
33	2	20	16	13	80.0	81.3	65.0
34	2	21	1	0	4.8	0.0	0.0
35	2	21	1	1	4.8	100.0	4.8
36	2	19	5	5	26.3	100.0	26.3
37	2	21	6	3	28.6	50.0	14.3
39	2	21	3	2	14.3	66.7	9.5
60	2	21	13	10	61.9	76.9	47.6
62	2	21	9	5	42.9	55.6	23.8
63	2	21	8	1	38.1	12.5	4.8
64	2	22	2	0	9.1	0.0	0.0
65	2	22	1	0	4.5	0.0	0.0
66	2	21	7	5	33.3	71.4	23.8
67	2	20	8	4	40.0	50.0	20.0
68	2	18	2	0	11.1	0.0	0.0
69	2	18	7	4	38.9	57.1	22.2
70	2	18	13	12	72.2	92.3	66.7
71	2	21	6	4	28.6	66.7	19.0
73	2	20	1	1	5.0	100.0	5.0
74	2	18	4	3	22.2	75.0	16.7
75	2	21	0	0	0.0		0.0
76	2	19	4	0	21.1	0.0	0.0
77	2	20	6	0	30.0	0.0	0.0
78	2	22	1	0	4.5	0.0	0.0
79	2	21	6	5	28.6	83.3	23.8
80	2	19	1	1	5.3	100.0	5.3
81	2	18	9	8	50.0	88.9	44.4
82	2	21	9	8	42.9	88.9	38.1
83	2	21	1	1	4.8	100.0	4.8
84	2	20	1	0	5.0	0.0	0.0
87	2	22	16	0	72.7	0.0	0.0
88	2	21	5	4	23.8	80.0	19.0
89	2	22	4	3	18.2	75.0	13.6
90	2	20	1	0	5.0	0.0	0.0
91	2	21	17	10	81.0	58.8	47.6
92	2	21	3	2	14.3	66.7	9.5
103	2	21	2	0	9.5	0.0	0.0
125	2	17	3	2	17.6	66.7	11.8
126	2	19	3	2	15.8	66.7	10.5
127	2	22	8	6	36.4	75.0	27.3
128	2	16	1	0	6.3	0.0	0.0
129	2	22	2	1	9.1	50.0	4.5
130	2	19	1	0	5.3	0.0	0.0

131	2	20	5	5	25.0	100.0	25.0
132	2	22	10	7	45.5	70.0	31.8
133	2	15	5	0	33.3	0.0	0.0
134	2	21	6	1	28.6	16.7	4.8
135	2	18	5	0	27.8	0.0	0.0
136	2	18	14	13	77.8	92.9	72.2
137	2	21	3	0	14.3	0.0	0.0
139	2	22	3	1	13.6	33.3	4.5
141	2	21	7	4	33.3	57.1	19.0
142	2	19	8	6	42.1	75.0	31.6
143	2	20	2	2	10.0	100.0	10.0
144	2	21	5	3	23.8	60.0	14.3
145	2	20	0	0	0.0		0.0
146	2	21	14	3	66.7	21.4	14.3
147	2	22	7	2	31.8	28.6	9.1
148	2	22	7	0	31.8	0.0	0.0
150	2	18	9	3	50.0	33.3	16.7
151	2	21	1	0	4.8	0.0	0.0
152	2	21	9	0	42.9	0.0	0.0
153	2	19	0	0	0.0		0.0
154	2	16	1	0	6.3	0.0	0.0
155	2	21	6	4	28.6	66.7	19.0
156	2	22	0	0	0.0		0.0
157	2	22	6	0	27.3	0.0	0.0
158	2	19	14	12	73.7	85.7	63.2
159	2	21	10	10	47.6	100.0	47.6
161	2	22	0	0	0.0		0.0
169	2	20	1	0	5.0	0.0	0.0
181	2	18	5	0	27.8	0.0	0.0
182	2	21	8	6	38.1	75.0	28.6
183	2	20	1	0	5.0	0.0	0.0
185	2	20	1	1	5.0	100.0	5.0
190	2	22	3	3	13.6	100.0	13.6
192	2	22	4	0	18.2	0.0	0.0
Total		**LPOT**	**TOT.L**	**TOT.R**	**% L**	**% R**	**% R/LPOT**
G1		866	25	2	2.9	8.0	0.9
G2		1577	413	225	26.2	54.5	14.3

Appendix 28. Rate of learning (L), participation points (P) and scores (PS) as well as late scores for participants (PPT) in the experimental group in study 3

PPT	L	PS	P	Late
31	10	31.8	7	9
32	6	27.3	6	6
33	16	95.5	21	3
34	1	36.4	8	0
35	1	40.9	9	1
36	5	40.9	9	30
37	6	63.6	14	12
39	3	4.5	1	13
60	13	118.2	26	6
62	9	63.6	14	50
63	8	-18.2	-4	49
64	2	-13.6	-3	33
65	1	72.7	16	5
66	7	81.8	18	12
67	8	109.1	24	28
68	2	-36.4	-8	0
69	7	68.2	15	59
70	13	50.0	11	1
71	6	81.8	18	1
73	1	13.6	3	11
74	4	4.5	1	17
75	0	31.8	7	12
76	4	50.0	11	12
77	6	86.4	19	8
78	1	9.1	2	4
79	6	95.5	21	3
80	1	81.8	18	25
81	9	22.7	5	185
82	9	27.3	6	0
83	1	63.6	14	9
84	1	27.3	6	1
87	16	59.1	13	59
88	5	90.9	20	5
89	4	9.1	2	50
90	1	104.5	23	23
91	17	118.2	26	7
92	3	13.6	3	98
103	2	-54.5	-12	7
125	3	18.2	4	1
126	3	-22.7	-5	12
127	8	27.3	6	2
128	1	-54.5	-12	0
129	2	109.1	24	25
130	1	50.0	11	47
131	5	-45.5	-10	11
132	10	18.2	4	8
133	5	22.7	5	186
134	6	-18.2	-4	100
135	5	50.0	11	95
136	14	113.6	25	8
137	3	-45.5	-10	7
139	3	-18.2	-4	337
141	7	4.5	1	7
142	8	68.2	15	261
143	2	63.6	14	6
144	5	81.8	18	16
145	0	72.7	16	6
146	14	27.3	6	3
147	7	-109.1	-24	0
148	7	18.2	4	31
150	9	90.9	20	270
151	1	-45.5	-10	20
152	9	0.0	0	0
153	0	-50.0	-11	16
154	1	-59.1	-13	0
155	6	127.3	28	2
156	0	50.0	11	16
157	6	0.0	0	24
158	14	0.0	0	6
159	10	50.0	11	1
161	0	-36.4	-8	54
169	1	-13.6	-3	10
181	5	-100.0	-22	0
182	8	31.8	7	262
183	1	81.8	18	180
185	1	18.2	4	12
190	3	-9.1	-2	1
192	4	4.5	1	6

Appendix 29. Initial questionnaire for study 3

Initial questionnaire

Grado (nombre de la titulación) *
Curso * ________
Edad * ________
Sexo *

- Hombre
- Mujer

¿Tienes acceso diario fácil a internet? *

- Sí
- No

¿Dónde te conectas a internet normalmente? *

- En casa
- En la universidad
- En el móvil

¿Cuánto tiempo pasas diariamente en internet? *
____

¿Utilizas frecuentemente las redes sociales? *

- Sí
- No

En caso afirmativo, ¿qué redes sociales usas?

- Facebook
- Twitter
- Tuenti
- LinkedIn
- Otro: ________

(En caso de respuesta afirmativa a la anterior) ¿Con qué frecuencia?

- 1 vez por semana
- 2/3 veces por semana
- 4/5 veces por semana
- 1 vez al día
- 2/3 veces al día
- 4 o más veces al día
- Otro: ________

(En caso de respuesta afirmativa a la anterior) ¿Cuánto tiempo pasas en las redes sociales?

- De 0 a 30 minutos
- De 30 minutos a una hora
- De una hora a dos horas
- Más de dos horas

¿Conoces Twitter? *

- Sí
- No

¿Sabes utilizar Twitter? *

- Sí
- No
- Creo que sí
- Creo que no

¿Utilizas Twitter habitualmente? *

- Sí
- No

(En caso afirmativo) ¿cuánto usas Twitter?

- Diariamente (varias veces al día)
- Diariamente (al menos una vez)
- Dos o tres veces a la semana
- Otro: ________

¿A qué hora sueles conectarte a internet o utilizar las redes sociales?

- Por la mañana
- Después de comer
- Por la tarde
- Por la noche
- Otro:

¿Tienes un teléfono móvil? *

- Sí
- No

(En caso afirmativo) ¿Es un smartphone?

- Sí
- No

(En caso afirmativo) ¿Tienes contratada una tarifa de datos para acceder a internet desde el teléfono?

- Sí
- No

¿Tienes alguna 'tablet' en casa? (tuya o de algún familiar) *

- Sí
- No
- Sí, de un familiar, pero puedo utilizarla
- Si, pero no la utilizo porque no es mía

¿Tienes un reproductor MP3/iPod/similar? *

- Sí
- No

(En caso afirmativo) ¿Escuchas podcasts o algún programa de radio similar desde tu MP3/reproductor de audio?

- Sí
- No
- Otro: ________

Si tienes alguna duda sobre el funcionamiento de los aparatos electrónicos que utilizas, ¿cómo la resuelves? *

- Leo el manual de instrucciones
- Intento averiguar cómo funciona por ti mismo
- Le pregunto a alguien
- Otro: ________

¿Cuántos años llevas estudiando/has estudiado inglés? * ________

¿Cómo definirías tu nivel de inglés? *

- Alto
- Medio-alto
- Medio-bajo
- Bajo

De todas las competencias de una lengua extranjera (gramática, vocabulario, pronunciación...) ¿cuál te resulta más difícil? *

____

¿Por qué? *

____

¿Utilizas algún tipo de herramienta tecnológica (ordenador, móvil, etc.) para ayudarte en tu aprendizaje de lengua extranjera? *

- Sí
- No

En caso de respuesta afirmativa en la anterior pregunta. Por favor, indica cuáles.

¿Alguna vez te has planteado si eres una persona sinestésica? *

- Sí
- No
- No lo sé
- No estoy seguro

¿Alguna vez has experimentado la sensación de que ciertas palabras te 'inspiran' ciertos colores? *

- Sí
- No
- No lo sé
- No estoy seguro

En caso afirmativo, ¿podrías dar algunos ejemplos de palabras que percibas asociadas a colores?

____

Appendix 30. Final questionnaire for study 3

Final questionnaire

*Obligatorio

¿Con qué frecuencia te has conectado a internet durante el estudio? *

- Una vez al día
- Dos veces al día
- Varias veces a la semana
- Otro: __________

Habitualmente, ¿cuánto tiempo pasas en Twitter? *

- Menos de 15 minutos al día
- Entre 15 y 30 minutos al día
- Entre 30 minutos y 1 hora
- Entre 1 y 2 horas
- Más de 2 horas

En tu uso de las redes sociales, ¿te consideras un usuario que participa de manera ACTIVA? (que escribe y comenta habitualmente) ¿o pasiva? (que le gusta leer información que otros publican pero no escribe mucho) *

- Que participa de manera activa
- Que participa de manera pasiva
- Otro: __________

Reflexionando sobre tu uso habitual de Twitter ¿Cuántos tweets envías habitualmente? *

- Menos de 1 tweet al día
- 1 o 2 tweets al día
- Entre 2 y 5 tweets al día
- Varios tweets por semana
- Varios tweets al mes
- Otro: __________

¿A qué cuenta estabas siguiendo durante el estudio? ¿Grupo 1 o 2? *

- Consejos generales (GRUPO 1)
- Consejos pronunciación (GRUPO 2)

¿Sospechaste en algún momento a lo largo del estudio que la meta era la pronunciación? *

- Sí
- No
- Otro: __________

En caso de que la anterior sea afirmativa, ¿estudiaste las palabras a conciencia para las entrevistas finales?

- Sí
- No
- Otro: __________

De ser afirmativa la anterior, ¿podrías indicarnos tu número de participante?
(Es únicamente para controlar los datos) Muchas gracias. __________

¿Cuál es tu postura sobre la siguiente afirmación? (¿es cierta para ti?) *
Twitter es una herramienta eficaz para la enseñanza

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

Del 1 al 5, 1 siendo 'muy poco útil' y 5 siendo 'muy útil' ¿qué nota le pondrías a Twitter para enseñar aspectos gramaticales? *

MUY POCO ÚTIL 1 2 3 4 5 MUY ÚTIL

Del 1 al 5, 1 siendo 'muy poco útil' y 5 siendo 'muy útil' ¿qué nota le pondrías a Twitter para enseñar falsos amigos? * Falsos amigos son palabras muy parecidas en inglés y en español, pero que en realidad no significan lo que parecen (ejemplo: 'carpet' no es carpeta, sino 'alfombra')

MUY POCO ÚTIL 1 2 3 4 5 MUY ÚTIL

Del 1 al 5, 1 siendo 'muy poco útil' y 5 siendo 'muy útil' ¿qué nota le pondrías a Twitter para enseñar pronunciación? *

MUY POCO ÚTIL 1 2 3 4 5 MUY ÚTIL

Del 1 al 5, 1 siendo 'muy poco útil' y 5 siendo 'muy útil' ¿qué nota le pondrías a Twitter para enseñar vocabulario? *

MUY POCO ÚTIL 1 2 3 4 5 MUY ÚTIL

¿Crees que te acordarás de los consejos recibidos a través de Twitter durante un tiempo? *

- Si
- No
- Otro: _________

¿Cuál es tu postura sobre la siguiente afirmación? *

Creo que el hecho de haber recibido los consejos a través de Twitter me hará recordarlos mejor que si me hubieran dado este consejo en clase

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Consideras que leer el consejo una vez es suficiente para consolidarlo? *

- Si
- No
- Otro: _________

¿Cuál es tu postura sobre la siguiente afirmación? *

Los consejos recibidos me han parecido útiles

Totalmente en desacuerdo 1 2 3 4 5 Totalmente de acuerdo

¿Has visto comprometida tu privacidad por compartir tu cuenta de Twitter con el profesor? *

- Si
- No
- Otro: _________

¿Cómo has leído los tweets? * (Puedes marcar varias opciones)

- Desde un ordenador
- Desde un teléfono móvil
- Desde una tableta
- Otro: _________

¿Dónde has leído los tweets? * (Puedes marcar varias opciones)

- En casa
- En la universidad
- En algún medio de transporte (bus, tranvía, etc.)
- En el trabajo
- Otro: _________

¿Podrías indicar qué sistema operativo utiliza tu teléfono móvil? *

- Android
- Windows
- Blackberry
- iOS (apple)
- Otro: _________

¿Dispones de alguna tablet en casa?

- Si
- No
- Si, pero no es mía.

En caso afirmativo ¿Podrías indicar el sistema operativo de la tablet?

- Android
- Windows
- Blackberry
- iOS (Apple)
- Otro: _________

OPCIONAL: Si tienes cualquier comentario adicional, por favor, hazlo aquí.

____

References

Abberton, E., & Fourcin, A. J. (1975). Visual feedback and the acquisition of intonation. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of Language Development* (pp. 157–165). New York, NY: Academic Press.

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching*, *3*, 113–122.

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly, 18,* 71–85.

Acton, W. (1997, February). Seven suggestions of highly successful pronunciation teaching. *Language Teacher Online, 21*(2). Retrieved from http://jalt-publications.org/old_tlt/files/97/feb/seven.html

Aist, G. (1999). Speech recognition in computer-assisted language learning. In K. Cameron (Ed.), *CALL: Media, design & applications* (pp. 165–181). Lisse, Holland: Swets & Zeitlinger.

Akahane-Yamada, R., Adachi, T., & Kawahara, H. (1997). Second language production training using spectrographic representations as feedback. *Journal of the Acoustical Society of Japan*, *18*, 341–343.

Akahane-Yamada, R., Adachi, T., Kawahara, H., Pruitt, J. S., & Mcdermott, E. (1998). Toward the optimization of computer-based second-language production training. *Proceedings of STiLL 98*, 111–114).

Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. *ICSLP 98, Proceedings of the 5th International Conference on Spoken Language Processing*. Rundle Mall, Australia: Casual Productions.

Albrechtsen, D., Henriksen, B., & Færch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning, 30,* 365–396. doi: 10.1111/j.1467-1770.1980.tb00324.x

Ally, M. (2007). Guest editorial. Mobile learning. *The international review of research in open and distance learning, 8*(2).

Amer, M. (2014). Language learners' usage of a mobile application for learning idioms and collocations. *CALICO Journal, 31*(3), 285–302.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369–406.

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth Publishers.

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes, 9*, 195–214.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20*(1), 51–62.

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5–22.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529–555.

Antenos-Conforti, E. (2009). Microblogging on Twitter: Social networking in intermediate Italian classes. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 59–90). San Marcos, TX: CALICO.

Apple Inc. (2014). *Oxford dictionary of English.* (Version 2.2.1 for Mac) [Computer software]

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science.* Cambridge, UK: Cambridge University Press

Atwell, E., Herron, D., Howarth, P., Morton, R., & Wick, H. (1999). *Pronunciation training: Requirements and solutions* (ISLE project report D1.4). Leeds, UK: University of Leeds. Retrieved from: http://www.cs.columbia.edu/~amaxwell/candidacy/l2learning/D14.pdf

Audacity Team. (2013). *Audacity.* (Version 2.0.5 for MAC). [Computer software]. Retrieved from: http://audacityteam.org/download/mac?lang=es

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford, UK: Oxford University Press.

Badin, P., Tarabalka, Y., Elisei, F., & Bailly, G. (2010). Can you 'read tongue movements'? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, *52*, 493–503.

Baker, A. (1992). *Introducing English pronunciation.* Cambridge, UK: Cambridge University Press.

Baker, A. (2006). *Ship or sheep? An intermediate pronunciation course.* Cambridge, UK: Cambridge University Press.

Baker, J. D. (2010). The impact of instructor immediacy and presence for online student affective learning, cognition, and motivation. *The Journal of Educators Online*, *7*(1), 1–30.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics, 44*(3), 231–250

Baran-Łucarz, M. (2014). Pronunciation anxiety and willingness to communicate in the FL classroom. *Proceedings of the international symposium on the acquisition of second language speech. Concordia working papers in applied linguistics*, *5*, 36–49.

Barreiro-Bilbao, S. C. (2013). Perception of natural and enhanced non-native contrasts in clear speech. *Onomázein, 27,* 207–219.

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 207–232). Baltimore, MD: York Press.

Beebe, L. (1978). Teaching pronunciation: Why we should be. *IDIOM, 9,* 2–3.

Bent, T., Bradlow, A., & Smith (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 331–347). Amsterdam, The Netherlands: John Benjamins.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication, 49*(10), 763–786.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171–203). Baltimore, MD: York Press.

Best, C. & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins Publishing Company.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech & Hearing Research*, *17*, 619–630.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly, 24*, 746–752.

Blattner, G., & Fiori, M. (2009). Facebook in the language classroom: Promises and possibilities. *Instructional Technology and Distance Learning*, *6*(1), 17–28.

Blattner, G., & Fiori, M. (2011). Virtual social network communities: An investigation of language learners' development of sociopragmatic awareness and multiliteracy skills. *CALICO Journal*, *29*(1), 24–43.

Blattner, G., & Lomicka, L. (2012). *Facebook*-ing and the social generation: A new era of language learning. *Alsic*, *15*(1), 1–22.

Blessing, S. B., Blessing, J. S., & Fleck, B. K. B. (2012). Using Twitter to reinforce classroom concepts. *Teaching of Psychology, 39*(4), 268–271.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer. (Version 5.5.17). [Computer program]. Retrieved from: http://www.praat.org/

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 273–304). Timonium, MD: York Press Inc.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NY: Lawrence Erlbaum

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30–50). Clevedon, UK: Multilingual Matters.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447–465.

Borau, K., Ullrich, C., Feng, J. J., & Shen, R. M. (2009). Microblogging for language learning: Using Twitter to train communicative and cultural competence. In M. Spaniol, Q. Li, R. Klamma & R. W. Lau (Eds.), *International Conference on Web-based Learning 2009* (pp. 78–87). Berlin, Germany: Spinger-Verlag.

Bradlow, A. R., Pisoni, D. B., Akahana-Yamada, R., & Tohkura, Y., (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19*, 51–61.

Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language & Speech, 24,* 207–221.

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y, & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication, 22*(2), 171–185.

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL, 16*(1), 103–113.

Brick, B. (2012). The role of social networking sites for language learning in UK higher education: The views of learners and practitioners. *International Journal of Computer Assisted Language Learning and Teaching, 2*(3), 35–53.

Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings* (pp. 211–224). London, UK: Routledge.

Brown, G. (1990). *Listening to spoken English* (2nd ed.). Harlow, UK: Longman Group UK Ltd.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System, 28,* 191–215.

Burleston, D. F. (2014). Improving intelligibility of non-native speech with computer-assisted phonological training. *IULC Working Papers, 7,* 1–18.

Busà, M. G. (2008). New perspectives in teaching pronunciation. In A. Baldry, M. Pavesi, C. Taylor Torsello, & C. Taylor (Eds.), *From DIDACTAS to ECOLINGUA. An ongoing research project on translation and corpus linguistics* (pp. 165–182). Trieste, Italia: Università degli Studi di Trieste.

Calvo Benzies, Y. J. (2013). 'Not enough time is devoted to pronunciation, I would like my textbook to include other types of pronunciation activities'. The views and perspectives of EFL fourth year secondary-school students on the teaching and learning of pronunciation. A survey-based study. In J. A. Mompean & J. Fouz-González (Eds.), *Proceedings of the third international conference on English pronunciation: issues & practices* (pp. 10–13). Murcia, Spain: Universidad de Murcia.

Cambridge University Press. (2011). *Clear Speech app.* (Version 1.01). [Mobile application software].

Carey, M. (2004). CALL visual feedback for pronunciation of vowels: Kay Sona-Speech. *CALICO Journal, 21*(3), 571–601.

Carlet, A., & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the 7th international symposium on the acquisition of second language speech. Concordia university working papers in applied linguistics*, *5*, 85–98.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training, research, and education* (pp. 87–136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation,* (pp. 87–100). Washington, DC: TESOL.

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal, 54*(7), 477–481.

Celce-Murcia, M. (1987). Teaching pronunciation as communication. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 5–12). Washington, DC: TESOL.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation. A reference for teachers of English to speakers of other languages* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code-mixing. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77–93). Clevedon, UK: Multilingual Matters.

Cenoz, J., & García Lecumberri, M. L. (1999a). The effect of training on the discrimination of English vowels. *International Review of Applied Linguistics, 37*(4), 261–275.

Cenoz, J., & García Lecumberri, M. L. (1999b). The acquisition of English pronunciation: learners' views. *International Journal of Applied Linguistics, 9*(1), 3–17.

Chapelle, C., & J. Jamieson. (2008). Tips for teaching with CALL: Practical approaches to computer-assisted language learning. White-Plains, NY: Pearson-Longman.

Chen, T. H., & Massaro, D. W. (2011). Evaluation of synthetic and natural Mandarin visual speech: Initial consonants, single vowels, and syllables. *Speech Communication, 53*, 955–972.

Cheung, W. S., & Hew, K. F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology, 25*(2), 153–183.

Chinnery, G. M. (2006). Emerging technologies: Going to the MALL: Mobile assisted language learning. *Language Learning & Technology*, *10*(1), 9–16. Retrieved from: http://llt.msu.edu/vol10num1/pdf/emerging.pdf

Chun, D. (1988). The neglected role of intonation in communicative competence and proficiency. *The Modern Language Journal, 72*(3), 295–303.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 61–77. Retrieved from: http://llt.msu.edu/vol2num1/pdf/article4.pdf

Chun, D. M. (2007). Technological advances in researching and teaching phonology. In M. C. Pennington (Ed.), *Phonology in context* (pp. 274–299). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Collins, B., & Mees, I. (2009). *Practical phonetics and phonology* (2nd ed.). London, UK: Routledge.

Colpaert, J. (2004). From courseware to coursewear? *Computer Assisted Language Learning, 17*(3-4), 261–266.

Coniam, D. (1999). Voice Recognition Software accuracy with second language speakers of English. *System, 27*(1), 49–64.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33,* 185–209.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161–170.

Cortés Moreno, M. (2000). Sobre la adquisición de la prosodia en lengua extranjera: Estado de la cuestión. *Didáctica (Lengua y Literatura), 12*, 91–119.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Cambridge, UK: Cambridge University Press. Retrieved from: http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, *18*(3), 53–70.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46–66.

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: socially constructed metalanguage and critical listening. *Language Awareness, 20*(3), 159–182.

Cranen, B., Weltens, B., de Bot, K., & van Rossum, N. (1984). An aid in language teaching: The visualization of pitch. *System, 12*(1), 25–29.

Cruttender, A. (2008). *Gimson's Pronunciation of English* (7th ed.). London, UK: Hodder Education.

Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Dordrecht, The Netherlands: Foris.

Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge, UK: Cambridge University Press.

Cutler, A. (1984). Stress and accent in language production and understanding. In D. Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm: studies in discourse phonology* (pp. 77–90). Berlin, Germany: de Gruyter.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp.183–196). Hillsdale, NJ: Lawrence Erlbaum.

Daniels, P. (2010). Developing iPhone apps for language learning. *JALT CALL Journal, 6*(2), 143–148.

Dayana Networks Ltd. (2015). *Voice record pro.* (Version 2.8.5). [Mobile software application]. Retrieved from: https://itunes.apple.com/es/app/voice-record-pro/id546983235?mt=8

de Jódar, O. (2008). La enseñanza de la pronunciación inglesa en el sistema educativo español (bachillerato). In R. Monroy & A. Sánchez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 231–235). Murcia, Spain: Universidad de Murcia.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*, 331–350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly, 16*, 71–77.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practical second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford, UK: Blackwell.

Dellar, H., & Walkley, A. (2005). *Innovations elementary. Coursebook*. Boston, MA: Heinle

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from Four L1s. *Studies in Second Language Acquisition, 20,* 1–16.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*, 379–397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly* 34, 592–603.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1997). Pronunciation instruction for 'fossilized' learners: Can it help? *Applied Language Learning, 8,* 217–235.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393–410.

Derwing, T. M., & Rossiter, M. (2002). ESL Learners' perceptions of their pronunciation needs and strategies. *System 30*, 155–166.

Derwing, T. M., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1–17.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). L2 fluency: Judgments on different tasks. *Language Learning*, *54*, 655–679.

Dickerson, W. (1987). Explicit rules and the developing interlanguage phonology. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 121–140). Dordrecht, The Netherlands: Foris.

Dickerson, W. B. (1994). Empowering students with predicative skills. In J. Morley (Ed), *Pronunciation pedagogy and theory: New directions, new views* (pp. 19–35). Bloomington, IN: TESOL.

Doughty, C., & Williams. J. (Eds.). (1998). *Focus on form in classroom second language acquisition.* Cambridge, UK: Cambridge University Press.

Dowd, A., Smith, J., & Wolfe, J. (1998). Learning to pronounce vowel sounds in a foreign language using acoustic measures of the vocal tract as feedback in real time. *Language and Speech*, *41*(1), 1–20.

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology, 13*(3), 66–86. Retrieved from: http://llt.msu.edu/vol13num3/ducatelomicka.pdf

Dudeney, G., & Hockly, N. (2007). *How to... teach English with technology.* Harlow, UK: Pearson Education Limited.

Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education*, *20*(2), 129–136.

Eddington, D. (2011). What are the contextual phonetic variants of /β ð ɣ/ in coloquial Spanish? *Probus, 23*(1), 1–19.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology, 2*(1), 54–73. Retrieved from: http://llt.msu.edu/vol2num1/pdf/article3.pdf

Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *JALT CALL Journal, 10*(1), 21–47.

Elliott, A. R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal, 79*, 356–371.

Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly, 27,* 91–113.

Ellis, R. (1997). *Second language acquisition.* Oxford, UK: Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching.* Cambridge, UK: Cambridge University Press.

English Computerized Learning Inc. (2010). *Pronunciation power.* (Version 2.0). [Mobile application software]. Retrieved from: https://itunes.apple.com/app/pronunciation-power/id368753108?mt=8

Engwall, O. (2008). Can audio-visual instructions help learners improve their articulation? An ultrasound study of short term changes. *Proceedings of INTERSPEECH 2008,* 2631–2634.

Engwall, O. (2012). Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning, 25*(1), 37–64.

Engwall, O., & Bälter, O. (2007). Pronunciation feedback from real and virtual language teachers. *Journal of Computer Assisted Language Learning, 20*(3), 235–262.

Engwall, O., Bälter, O., Öster, A. M., & Kjellström, H. (2006). Feedback management in the pronunciation system ARTUR. *Proceedings of the ACM SIGHCI conference on human factors in computing systems*, 231–234. Retrieved from: ftp://ftp.nada.kth.se/pub/documents/IPLab/TechReports/HCI-16.pdf

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication, 51*(10), 832–844.

Fagel, S., & Madany, K. (2008). A 3-D virtual head as a tool for speech therapy for children. *Proceedings of INTERSPEECH 2008*, 2643–2646.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intellibility and irritation. *Language Learning, 37,* 313–326.

Fayer, J., & Krasinski, E. (1995). Perception of hesitation in nonnative speech. *Bilingual Review, 20*, 114–121.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399–423.

Firth, S. (1987). Developing self-correcting and self-monitoring strategies. *TESL Talk, 17*(1), 148–152.

Firth, S. (1992). Pronunciation syllabus design: a question of focus. In P. Avery & S. Ehrlich (Eds.), *TeachingAmerican English pronunciation* (pp. 173–183). Oxford, UK: Oxford University Press.

Flege, J. E. (1987). Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 9–39). Dordrecht, The Netherlands: Foris.

Flege, J. E. (1989). Using visual information to train foreign language vowel production. *Language Learning*, *38*, 365–407.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Hueber & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 249–289). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins

Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233–277). Timonium, MD: York Press Inc.

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol, UK: Multilingual Matters.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America, 83*, 729–740.

Flege, J. E., & Hammond, R. M. (1982). Mimicry of non-distinctive phonetic difference between language varieties. *Studies in Second Language Acquisition, 5,*1–17.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1–34.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97,* 3125–3134.

Fouz-González, J. (2012). Can Apple's iPhone help to improve English pronunciation autonomously? State of the app. In L. Bradley & S. Thouësny (Eds.), *CALL: Using, learning, knowing, EUROCALL conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings* (pp. 81–87). Dublin, UK: Research-publishing.net.

Fouz-González, J. (2013, May). How can mobile phone apps enhance the learning of pronunciation? Reasons to 'go mobile'. Paper presented at the EPIP3 3rd International Conference on English Pronunciation: Issues and Practices, Murcia, Spain.

Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In J. A. Mompean & J. Fouz-González. *Investigating English pronunciation: Trends and directions* (pp. 314–342). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Fouz-González, J., & Mompean, J. A. (2012). Twitter and English pronunciation teaching. In Pixel (Ed.) *ICT for Language Learning 5th edition. Conference Proceedings* (pp. 341–344). Florence, Italy: Libreriauniversitaria.it.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97,* 2540–2551.

Fraser, H. (2001). *Teaching Pronunciation: A handbook for teachers and trainers*. Sydney, Australia: TAFE Access Division.

Fraser, H. (2006). Phonological concepts and concept formation: metatheory, theory and application. *International Journal of English Studies, 6*(2), 55–75. Retrieved from: http://revistas.um.es/ijes/article/view/48801

Gass, S. M. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

García Cabrero, J. C. (2002). Third generation telephony: New technological support for computer assisted language learning. *International Journal of English Studies, 2*(1), 167–178. Retrieved from: http://revistas.um.es/ijes/article/view/48591/46511

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115–135). Clevedon, UK: Multilingual Matters.

Gick, B., Bernhardt, M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 309–322). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Gilbert, J. B. (1994). Intonation: A navigation guide for the listener. In J. Morley (Ed.), *Pronunciation pedagogy and theory* (pp. 36–48). Bloomington, IN: TESOL.

Gilbert, J. (2008). Teaching pronunciation using the prosody pyramid. New York, NY: Cambridge University Press.

Gilbert, J. (2010). Pronunciation as orphan: What can be done? *As We Speak: Newsletter of TESOL speaking, pronunciation, listening interest section, 7(2).*

Guion & pederson 2007.

Godwin-Jones, R. (2005a). Emerging technologies. Skype and podcasting: Disruptive technologies for language learning. *Language Learning & Technology*, *9*(3), 9–12. Retrieved from: http://llt.msu.edu/vol9num3/emerging/

Godwin-Jones, R. (2005b). Emerging technologies: Messaging, gaming, peer-to-peer sharing: Language learning strategies & tools for the millenial generation. *Language Learning & Technology*, *9*(1), 17–22. Retrieved from: http://llt.msu.edu/vol9num1/emerging/

Godwin-Jones, R. (2009). Emerging technologies: Speech tools and technologies. *Language Learning & Technology* *13*(3), 4–11. Retrieved from: http://llt.msu.edu/vol13num3/emerging.pdf

Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *Language Learning & Technology,* *15*(2), 2–11. Retrieved from: http://llt.msu.edu/issues/june2011/emerging.pdf

Gómez, P., Álvarez, A., Martínez, R., □□□□□□□ Bobadilla, J., Bernal, J., Rodellar, V., & Nieto, V. (2008). Applications of formant detection in language learning. In V. M. Holland & F. P. Fisher (Eds.) *The path of speech technologies in computer assisted language learning,* (pp. 44–65). New York, NY: Routledge.

González-Bueno, M. (1997). The effects of formal instruction on Spanish stop consonants. In W. Glass & A. T. Pérez-Leroux (Eds.), *Contemporary perspectives on the acquisition of Spanish: Production, processing and comprehension* (Vol. 2, pp. 57–75). Somerville, MA: Cascadilla.

Gordon, P. C., Keyes, L., & Yung, Y. F. (2001). Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli. *Perception and Psychophysics*, *63*, 746–758.

Goto, H. (1971). Auditory perception by normal Japanese subjects of the sounds “L” and “R”. *Neuropsychologia, 9,* 317–323.

Guiora, A. Z., Acton, W. R., Erard, R., & Strickland, F. W. (1980). The effects of benzodiazepine (Valium) on permeability of language ego boundaries. *Language Learning, 30,* 351–364.

Guiora, A. Z., Beit-Hallahmi B., Brannon, R. C. L., Dull, C. Y., & Scovel, T. (1972). The effects of experimentally induced changes into ego states on pronunciation ability in a second language: an exploratory study. *Comprehensive Psychiatry, 13*, 421–428.

Gutiérrez-Díez, F. (2001). The acquisition of English syllable timing by native speakers learners of English. An empirical study. *International Journal of English Studies, 1*(1), 93–114. Retrieved from: http://revistas.um.es/ijes/article/view/47631

Gynan, N. (1985). Comprehension, irritation and error hierarchies. *Hispania, 68*, 160–165.

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201–223.

Han, Z., Park, E., & Combs, C. (2008). Textual enhancement of input: Issues and possibility. *Applied Linguistics, 29*(4), 597–618.

Hancock, M. (2003). *English pronunciation in use*. Cambridge, UK: Cambridge University Press.

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication, 51*(10), 906–919.

Handley, Z., & Hamel, M. J. (2005). Establishing a methodology for benchmarking speech synthesis for computer-assisted language learning (CALL). *Language Learning & Technology*, *9*(3), 99–119. Retrieved from: http://llt.msu.edu/vol9num3/pdf/handley.pdf

Hardison, D. M. (1999). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning, 49,* 213–283.

Hardison, D. M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*, 495–522.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34–52. Retrieved from: http://www.llt.msu.edu/vol8num1/pdf/hardison.pdf

Hardison, D. M. (2005). Contextualised computer-based L2 prosody training: evaluating the effects of discourse context and video input. *CALICO Journal, 22*(2), 175–190.

Hardison. (2007). The visual element in phonological perception and learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 135–158). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Hardisty, D., & Windeatt, S. (1989). *Computer-assisted language learning.* Oxford, UK: Oxford University Press.

Harrison, R. (2013). Profiles in social networking sites for language learning – *Livemocha* revisited. In M.-N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 100–116). Basingstoke, UK: Palgrave Macmillan.

Hazan, V., & Li, E. (2008). The effect of auditory and visual degradation on audiovisual perception of native and non-native speakers. *Proceedings of INTERSPEECH*, 1191–1194.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005) Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*(3), 360–378.

Hazan, V., & Simpson, A. (1998). The effect of cue-enhancement on consonant perception by non-native listeners: preliminary results. *Proceedings of STiLL 98*, 119–122.

Hazan, V., & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: talker and listener effects. *Language and Speech 43*(3), 273–294.

Henderson, A., Curnick, L., Frost, D., Kautzsch, A., Kirkova-Naskova, A., Levey, D., Tergujeff, E., & Waniek-Klimczak, E. (2015). The English pronunciation teaching in Europe survey: Factors inside and outside the classroom. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Current trends and directions* (pp. 260–292). Basingstoke, UK/New York, NY: Palgrave Macmillan

Herron, D., Menzel, W., Atwell, E., Bisiani, R., Daneluzzi, F., Morton, R., & Schmidt, J. A. (1999). Automatic localization and diagnosis of pronunciation errors for second-language learners of English. *Eurospeech'99. 6th European Conference on Speech Communication and Technology*, 855–858.

Hew, S. H., & Ohki, M. (2001). A study on the effectiveness and usefulness of animated graphical annotation in Japanese CALL. *ReCALL, 13*(2), 245–260.

Hew, S. H., & Ohki, M. (2004). Effect of animated graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal, 21*(2), 397–420.

High Criteria Inc. (2013). Total Recorder. Professional edition. (Version 8.5) [Computer software]. Retrieved from: http://www.totalrecorder.com/downloads.htm

Hincks, R. (2001). Using speech recognition to evaluate skills in spoken English. *Working Papers, 49*, 58–61.

Hincks, R. (2002). Speech synthesis for teaching lexical stress. *Proceedings of Fonetik, TMH-QPSR*, *44*(1), 153–156.

Hincks, R. (2003). Speech technologies for pronunciation, feedback and evaluation. *ReCALL, 15*(1), 3–20.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System, 33*(4), 575–591.

Hincks, R. (2015). Technology and learning pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 505–519). Malden, NY: Wiley- Blackwell.

Hincks, R., & Edlund, J. (2009). promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology, 13*(3), 32–50. Retrieved from: http://llt.msu.edu/vol13num3/hincksedlund.pdf

Hinofotis, F., & Bailey, K. (1980). American undergraduate reactions to the communication skills for foreign teaching assistants. In J. Fisher, M. Clarke, & J. Schacter (Eds.), *On TESOL '80: Building bridges* (pp. 120–133). Washington, DC: TESOL.

Hoffman, R. (1996). Computer networks: Webs of communication for language teaching. In M. Pennington (Ed.), *The power of CALL* (pp. 55–78). Houston, TX: Athelstan publications.

Horgues, C., & Scheuer, S. (2015). Why some things are better done in tandem. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Current trends and directions* (pp. 47–82). Basingstoke, UK/New York, NY: Palgrave Macmillan

Hornby, A. S. (2010). *Oxford advanced learner's dictionary* (8th ed.). Oxford, UK: Oxford University Press.

Horwitz, E. K., Horwitz, M., & Cope, J. A. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70,* 125–132.

Hualde, J. I. (2014). *Los sonidos del español.* Cambridge, UK: Cambridge University Press.

Huang, X., Acero, A., & Hon, H. W. (2004). *Spoken language processing: A guide to theory, algorithm and system development*. Engle-wood Cliffs, NJ: Prentice Hall.

Information Technology Ltd. (2014). *King of vowels.* (Version 2.0.1). [Mobile application software].

Ioup, G. (1995). Evaluating the need for input enhancement in post-critical period language acquisition. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 95–123). Clevedon, UK: Multilingual Matters

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73–98.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, *27*(1), 1–12.

Iverson, P., Hazan, V., & Banister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118,* 3267–3278.

James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *International Review of Applied Linguistics, 14*(3), 227–243.

James, E. (1977). The acquisition of second language intonation using a visualizer. *Canadian Modern Language Review, 33*(4), 503–506.

Jamieson, D., & Morosan, D. (1986). Training non-native speech contrasts in adults: Acquistion of the English /ð/-/θ/ contrast by francophones. *Perceptual Psychophysics*, *40*, 205–215.

Jamieson J., Chapelle C., & Preiss S. (2005). CALL evaluation by developers, a teacher, and students. *CALICO Journal, 23*(1), 93–125.

Jansen, B. J., & Zhang, M. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology, 60*(11), 2169–2188.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. *Proceedings of the Joint 9th WEBKDD & 1st SNA-KDD Workshop 2007*, 56–65.

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford, UK: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically-based, empirically-researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83–103.

Jensen, J. (2011). Readers' forum: Phoneme acquisition: Infants and second language learners. *The Language Teacher Online, 35*(6) 24–28.

Jones, D., Roach, P., Setter, J., & Esling, J. (2011). *Cambridge English pronouncing dictionary* (18th ed). Cambridge, UK: Cambridge University Press.

Jones, R. H. (1997). Beyond 'Listen and Repeat': Pronunciation teaching materials and theories of second language acquisition. *System*, *25*, 103–112.

Junco, R., Elavsky, M. C., & Heiberger, G. (2013). Putting twitter to the test: Assessing outcomes for student collaboration, engagement and success. *British Journal of Educational Technology, 44*(2), 273–287.

Junco, R., Heibergert, G., & Loken, E. (2010). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning, 27*, 119–132.

Kang, M., Kashiwagi, H., Treviranus, J., & Kaburagi, M. (2008). Synthetic speech in foreign language learning: An evaluation by learners. *International Journal of Speech Technology, 11*(2), 97–106.

Kassens-Noor, E. (2012). Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets. *Active Learning in Higher Education, 13*(1), 9–21.

Kazuya Kamioka. (2011a). *English Ear Game*. (Version 1.7.2). [Mobile application software]. Retrieved from: https://itunes.apple.com/us/app/id433419223?mt=8&ign-mpt=uo%3D4

Kazuya Kamioka. (2011b). *English Ear Game 2*. (Version 1.7.2). [Mobile application software]. Retrieved from: https://itunes.apple.com/us/app/id448921428?mt=8&ign-mpt=uo%3D4

Kelly*,* G. (2000). *How to teach pronunciation*. Harlow, UK: Longman.

Kennedy, A. S. (2015). Non-native accent and listener perceptions of grammaticality. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Current trends and directions* (pp. 131–146). Basingstoke, UK/NewYork, NY: Palgrave Macmillan.

Kennedy, C., & Levy, M. (2008). L'italiano al telefonino: Using SMS to support beginners' language learning. *ReCALL, 20*(3), 315–330.

Kenning, M-M. (2007). *ICT and language learning. From print to the mobile phone*. Basingstoke, UK/New York, NY: Palgrave Macmillan.

Kenning M-M., & Kenning M. J. (1990) *Computers and language learning: current theory and practice*. New York, NY: Ellis Horwood.

Kenworthy, J. (1987). *Teaching English pronunciation*. New York, NY: Longman.

Kim, D., Rueckert, D., Kim, D-J., & Seo, D. (2013). Students' perceptions and experiences of mobile learning. *Language Learning & Technology, 17*(3), 52–73. Retrieved from: http://llt.msu.edu/issues/october2013/kimetal.pdf

Kim, H., & Kwon, Y. (2012). Exploring smartphone applications for effective mobile-assisted language learning. *Multimedia-Assisted Language Learning, 15*(1), 31–57.

Kim, H., Park, S-M., & Baek, S-H. (2011). Twitter and implications for its use in EFL learning. *Multimedia Assisted Language Learning, 14*(2), 113–137.

Kim, H-S. (2013). Emerging mobile apps to improve English listening skills. *Multimedia-assisted language learning, 16*(2), 11–30.

Kim, I-S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society*, *9*(1), 322–334.

King, R. D. (1967). Functional load and sound change. *Language, 43*, 831–852.

Knight, R-A. (2010). Sounds for study: Speech and language therapy students' use and perception of exercise podcasts for phonetics. *International Journal of Teaching and Learning in Higher Education, 22*(3), 269–276.

Kondo, M., Ishikawa, Y., Smith, C., Sakamoto, K., Shimomura, H., & Wada, N. (2012). Mobile assisted language learning in university EFL courses in Japan: developing attitudes and skills for self-regulated learning. *ReCALL*, *24*(2), 169–187.

Krashen, S. (1979). Adult second language acquisition as post-critical period learning. *ITL: Review of Applied Linguistics, 43*, 39–52.

Krashen, S. (1981). Aptitude and attitude in relation to second language acquisition and learning. In C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 155–175). Rowley, MA: Newbury House.

Krashen, S. (1982). Principles and practice in second language acquisition. Oxford, UK: Pergamon Institute.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Oxford, UK: Pergamon Press.

Kröger, B., Graf-Borttscheller, V., & Lowit, A. (2008). Two- and three-dimensional visual articulatory models for pronunciation training and for treatment of speech disorders. *Proceedings of INTERSPEECH 2008*, 2639–2642.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnetic effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics 50*, 93–107.

Kuhl, P., & Iverson, P. (1995). Linguistic experience and the 'perceptual magnet effect'. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 121–154). Timonium, MD: York Press Inc.

Kukulska-Hulme, A. (2005). Mobile usability and user experience. In A. Kukulska-Hulme & J. Traxler (Eds.), *Mobile learning: A handbook for educators and trainers* (pp. 45–56). London, UK: Routledge.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston, MA: Wadsworth Cengage.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning, 12*(2), 137–156.

Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26,* 227–247.

Lamy, M-N., & Zourou, K. (2013). Introduction. In M-N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 1–7). Basingstoke, UK: Palgrave Macmillan.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford, UK: Oxford University Press.

Leather, J. (2003). Phonological acquisition in multilingualism. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 23–58). Clevedon, UK: Multilingual Matters.

Leather, J. H., & James, A. R. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition, 13,* 305–341.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York, NY: John Wiley & Sons, Inc.

Léon, P., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *Modern Language Journal, 56*, 139–144.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. In S. Pemberton (Ed.), *Human factors in computing systems* (pp. 359–366). Atlanta, Georgia: Association for Computing Machinery.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093–1096.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 369–377.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27,* 184–202.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*(4), 505–524.

Levy, M. (1997a). *Computer-assisted language learning. Context and conceptualization.* Oxford, UK: Clarendon Press.

Levy, M. (1997b). Theory-driven CALL and the development process. *Computer-Assisted Language Learning, 10*(1), 41–56.

Levy, M., & Kennedy, C. (2005). Learning Italian via mobile SMS. In A. Kukulska-Hulme & J. Traxler (Eds.), *Mobile learning: A handbook for educators and trainers.* London, UK: Taylor and Francis.

Levy, M., & Stockwell, G. (2006). *CALL Dimensions. Options and issues in computer-assisted language learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology, 17*(3), 135–156. Retrieved from: http://llt.msu.edu/issues/october2013/lihegelheimer.pdf

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25. Retrieved from: http://www.equinoxpub.com/journals/index.php/CALICO/article/view/22942

Liu, M., Evans, M., Horwitz, E., Lee, S., McCroy, M., Park, J-B., & Parrish, C. (2013). A study of the use of social network sites for language learning by university ESL students. In M-N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 137–157). Basingstoke, UK: Palgrave Macmillan.

Liu, M. L., Moore, Z., GrahaAm, L., & Lee, S. (2003). A look at the research on computer-based technology use in second language learning: A review of the literature from 1990-2000. *Journal of Research on Technology in Education, 34*(3), 250–273.

Liu, Y., Massaro, D. M., Chen, T. H., Chan, H. L., & Perfetti, C. (2007). Using visual speech for training Chinese pronunciation: an in-vivo experiment. In *Proceedings ISCA ITRW* SLaTE, Farmington, PA.

Lively, S., Logan J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environments and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In ICPhS 1995. Proceedings of the 13th International Congress of Phonetic Sciences. Vol 4. (pp. 92–99). Stockholm, Sweden, August 13-19, 1995.

Llisterri, J. (2001). Enseñanza de la pronunciación, corrección fonética y nuevas tecnologías. *Es Espasa, Revista de Profesores,* 28 de noviembre de 2001.

Llisterri, J. (2007). La enseñanza de la pronunciación asistida por ordenador. In Actas del XXIV Congreso Internacional de AESLA. Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva: Perspectivas aplicadas entre disciplinas. [CD-ROM] (pp. 91-120). Madrid: Universidad Nacional de Educación a Distancia - Asociación Española de Lingüística Aplicada. Retrieved from: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_06_Pronunciacion_Tecnologias.pdf

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874–886.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 351–378). Baltimore, MD: York Press.

Lomicka, L., & Lord, G. (2009). Introduction to social networking, collaboration, and web 2.0 tools. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 1–11). San Marcos, TX: CALICO.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Eds), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam, The Netherlands: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania, 88*(3), 557–567.

Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals, 41,* 374–389.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*(3), 488–503.

Lord, G., & Lomicka, L. (2014). Twitter as a tool to promote community among language teachers. *Journal of Technology and Teacher Education, 22*(2), 187–212.

Lowe, B., & Laffey, D. (2011). Is Twitter for the birds? Using Twitter to enhance student learning in a marketing course. *Journal of Marketing Education, 33*(2), 183–192.

Lys, F. (2013). The development of advanced learner oral proficiency using iPads. *Language Learning & Technology*, *17*(3), 94–116. Retrieved from http://llt.msu.edu/issues/october2013/lys.pdf

MacCarthy, P. (1978). *The teaching of pronunciation.* Cambridge, UK: Cambridge University Press.

MacDonald, S. (2002). Pronunciation – views and practices of reluctant teachers. *Prospect, 17*(3), 3–18.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253–257.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning, 44,* 75–100.

Macmillan Publishers Ltd. (2011). *Sounds*: *The pronunciation app.* (Version 3.0). [Mobile application software]. Retrieved from: https://itunes.apple.com/gb/app/sounds-the-pronunciation-app/id442713833?mt=8

Mairs, J. L. (1989). Stress assignment in interlanguage phonology: An analysis of the stress system of Spanish speakers learning English. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on SLA* (pp. 260–283). Cambridge, UK: Cambridge University Press.

Major, R. (1987). The natural phonology of second language acquisition. In A. James & J. Leather (Eds.). *Sound patterns in second language acquisition* (pp. 207–224). Dordrecht, The Netherlands: Foris.

Major, R., & Kim, E. (1999). The similarity differential rate hypothesis. In J. Leather (Ed.), *Phonological issues in language learning* (pp. 151–184). Oxford, UK: Blackwell.

Martínez-Celdrán, E., Fernández-Planas, A. M. & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association, 33*(2), 255–259.

Massaro, D. W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Massaro, D. W. (2006). The psychology and technology of talking heads: applications in language learning. In O. Bernsen, L. Dybkjaer, & J. van Kuppevelt (Eds.), *Natural, intelligent and effective interaction in multimodal dialogue systems* (pp. 183–214). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Massaro, D., Bigler, S., Chen, T., Perlman, M., & Ouni, S. (2008). Pronunciation training: The role of eye and ear. In *Interspeech* (pp. 2623–2626). Geneva, Switzerland: International Speech Communication Association (ISCA).

Massaro, D. W., Cohen, M. M., Tabain, M., Beskow, J., & Clark, R. (2012). Animated speech: Research progress and applications. In G. Bailly, P. Perrier, & E. Vatikiotis-Bateson (Eds.) *Audiovisual speech processing* (pp. 309–345). Cambridge, UK: Cambridge University Press.

Massaro, D. W., & Light, J. (2003). Read My Tongue Movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In *Proceedings of Eurospeech*

(Interspeech), 8th European Conference on Speech Communication and Technology (pp. 2249-2252). Geneva, Switzerland.

Massaro, D. W., & Light, J. (2004). Improving the vocabulary of children with hearing loss. *Volta Review, 104*(3), 141–174.

Massaro, D. W., Liu, Y., Chen, T. H., & Perfetti, C. A. (2006). A multilingual embodied conversational agent for tutoring speech and language learning. In *Proceedings of the 9th International Conference on Spoken Language Processing* (Interspeech 2006 – ICSLP), (pp. 825–828) Pittsburgh, PA.

Mayer, R. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19.

Mayer, R., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.

Mayfield Tomokiyo, L., Wang, L., & Eskenazi, M. (2000) An empirical study of the effectiveness of speech-recognition-based pronunciation tutoring. *Proceedings of the 6th International Conference on Speech and Language Processing*, Beijing, China, 2000, 677-680.

Mbah, E. (2014). Effect of podcast on the Igbo learners of English phonetics. *International Journal of Humanities and Social Sciences Research, 1*(1), 24–42.

McBride, K. (2009a). Podcasts and second language learning: Promoting listening comprehension and intercultural competence. In L. B. Abraham & L. Williams (Eds.), *Electronic discourse in language learning and language teaching*. Amsterdam, The Netherlands: John Benjamins Publishing Co.

McBride, K. (2009b). Social networking sites in foreign language classes: Opportunities for re-creation. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 35–58). San Marcos, TX: CALICO.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Menzel, W., Herron, D., Morton, R., Pezzotta, D., Bonaventura, P., & Howarth, P. (2001). Interactive pronunciation training, *ReCALL*, *13*(1), 67–78.

Molholt, G. (1988). Computer-Assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*(1), 91–111.

Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities, 24*, 81–92.

Molholt, G., & Hwu, F. (2008). Visualization of speech patterns for language learning. In V. M. Holland & F. P. Fisher (Eds.), *The path of speech technologies in computer assisted language learning* (pp. 91–122). New York, NY: Routledge.

Mompean, J. A. (2001). A comparison between English and Spanish subjects' typicality ratings in phoneme categories: A first report. *International Journal of English Studies, 1*(1), 115–156. Retrieved from: http://revistas.um.es/ijes/article/view/47641/45661

Mompean, J. A. (2002). The categorisation of the sounds of English: Experimental evidence in phonology. (Unpublished doctoral dissertation). Murcia, Spain: Universidad de Murcia.

Mompean, J. A. (2004a). Category overlap and neutralization: The importance of speakers' classifications in phonology. *Cognitive Linguistics, 15*(4), 429–469.

Mompean, J. A. (2004b). Options and criteria for the choice of an English pronunciation model in Spain. In J. M. Oro (Ed.), *Linguistic perspectives from the classroom* (pp. 1043–1059). Santiago de Compostela, Spain: Servizo de Publicacións.

Mompeán, J. A. (2005). Taking advantage of phonetic symbols in the foreign language classroom. *Proceedings of the Phonetics Teaching and Learning Conference 2005*, 50–54.

Mompean, J. A. & Fouz González, J. (in press, 2016). Twitter-based EFL pronunciation instruction. *Language Learning & Technology, 20*(1).

Mompeán-Guillamón, P. (2014). Vowel–colour symbolism in English and Arabic: A comparative study. *Miscelánea*, *47*, 31–52.

Mompeán-Guillamón, P. (2015). Sound Symbolism and Pronunciation Teaching: a Preliminary Study. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 243–259). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Monroy-Casas, R. (1980). *Aspectos fonéticos de las vocales españolas.* Madrid: Sociedad General Española de Librería (SGEL).

Monroy, R. (2001). Profiling the phonological processes shaping the fossilised IL of adult Spanish learners of English. Some theoretical implications. *International Journal of*

English Studies, 1, 157–217. Retrieved from: http://revistas.um.es/ijes/article/view/47661

Mork, C. (2009). Using Twitter in EFL education. *JALT CALL Journal, 5*(3), 41–56.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, *25*(3), 481–520.

Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.) *Pronunciation, pedagogy and the theory: New views, new directions* (pp. 64–91). Bloomington, IN: TESOL.

Morton, H., & Jack, M. (2005). Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning*, *18*, 171–191.

Morton H., Gunson, N., & Jack, M. (2012). Interactive language learning through speech-enabled virtual scenarios. *Advances in Human-Computer Interaction, 12*, 1–14.

Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology, 13*(2), 29–47. Retrieved from: http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition, 21*, 81–108.

Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten (Eds), *Input matters in SLA* (pp. 159–174). New York, NY: Multilingual Matters

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal, 20*(2), 38–51

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*, 520–531.

Munro, M. J. & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching, 44*, 316–327.

Murphy, M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, *42*, 258–269.

Murray, G. (1999). Autonomy and language learning in a simulated environment. *System,* 27, 295–308.

Nelson, T., & Oliver, W. (1999). Murder on the Internet. *CALICO Journal, 17*(1), 101–114.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441–467.

Neri, A., Cucchiarini, C., & Strik, H. (2008) The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch. *ReCALL*, *20*(2), 225–243.

Nunan, D. (1989). *Designing tasks for the communicative classroom.* Cambridge, UK: Cambridge University Press.

O'Brien, M. (2006). Teaching pronunciation and intonation with computer technology. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 127–148). San Marcos, TX: CALICO.

O'Bryan, A., & Hegelheimer, V. (2007). Integrating CALL into the classroom: the role of podcasting in an ESL listening strategies course. *ReCALL, 19*(2), 162–180.

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology, 18*(3), 173–192. Retrieved from: http://llt.msu.edu/issues/october2014/olson.pdf

Ortega Llebaria, M. (2004). Interplay between phonetic inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In F. L Timothy (Ed.), *Laboratory approaches to Spanish phonology* (pp. 237–253). Berlin, Germany: Mouton de Gruyter.

Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001) Auditory-visual L2 speech perception: effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Speech, Hearing and Language: UCL Work in Progress*, *13*, 39–41.

Ota, F. (2011). A study of social networking sites for learners of Japanese. *New Voices*, *4*, 144–167.

Ouni, S., Cohen, M. M., Young, K., & Jesse, A. (2003). Internationalization of a talking head. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of 15th International Congress of Phonetics Sciences* (pp. 2569–2572). Barcelona, Spain: Casual Productions.

Oxenden, C., Latham-Koenig, C., & Seligson, P. (2004). *New English file. Elementary student's book*. Oxford, UK: Oxford University Press.

Oxford University Press. (2012). *English File Pronunciation.* (Version 1.1). [Mobile application software]. Retrieved from: https://itunes.apple.com/es/app/english-file-pronunciation/id520767531?mt=8

Paulston, C. B. (1972). Structural pattern drills: A classification. In H. Allen & R. Campbell (Eds.) *Teaching English as a second language* (pp. 129–138). New York, NY: McGraw-Hill.

Pennington, M. C. (1989). Teaching pronunciation from the top down. *RELC Journal*, *20*(1), 20–38.

Pennington, M. C. (1994). Recent research in L2 phonology: Implications for practice. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New ways, new directions* (pp. 92–107). Bloomington, IN: TESOL.

Pennington, M. C. (1996). The power of the computer in language education. In M.C. Pennington (Ed.), *The Power of CALL*. Houston, TX: Athelstan publications.

Pennington, M. C. (1998). The teachability of pronunciation in adulthood: A reconsideration. *International Review of Applied Linguistics, 36*, 323–41.

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: promise, limitations, directions. *Computer-Assisted Language Learning, 12*(5), 427–440.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84,* 372–389.

Pennington, M. C., & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In M. C. Pennington (Ed.) *The power of CALL* (pp. 153–189). Houston, TX: Athelstan.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterlly, 20,* 207–225.

PhoTransEdit. (2013). *PhoTransEdit.* (Version 1.7f). [Computer software]. Retrieved from: http://www.photransedit.com/Default.aspx

Pi-Hua, T. (2006). Bridging pedagogy and technology: User evaluation of pronunciation oriented CALL software. *Australasian Journal of Educational Technology*, *22*(3), 375–397.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, *35*(2), 233–255.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29,* 191–215.

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *Journal of the Acoustical Society of America*, *70*, S98.

Pisoni, D. B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, *1*, 10–23.

Prensky, M. (2001). Digital natives, digital immigrants Part 1. *On the Horizon 9*(5), 1–6.

Probst, K., Ke, Y., & Eskenazi, M. (2002). Enhancing foreign language tutors – in search of the golden speaker. *Speech Communication*, *37*(3–4), 161–173.

Promnitz-Hayashi, L. (2011). A learning success story using Facebook. *Studies in Self-Access Learning Journal, 2*(4), 309–316.

Pruitt, J. S., Kawahara, H., Akahane-Yamada, R., & Kubo, R. (1998). Methods of enhancing speech stimuli for perceptual training: Exagerated articulation, context truncation, and 'STRAIGHT' re-synthesis. *Proceedings of STiLL 1998,* 107–110. Retrieved from: http://www.isca-speech.org/archive_open/archive_papers/still98/stl8_107.pdf

Psyentific Mind Inc. (2012). iBaldi. (Version 2.1) [Mobile application software]. Retrieved from: https://itunes.apple.com/es/app/ibaldi/id504464546?mt=8

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30,* 271–287.

Quillis, A. (2012). *Principios de fonología y fonética españolas.* Madrid, Spain: Arco Libros.

Ramírez-Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness, 15*(3), 141–159.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). New York, NY: Cambridge University Press.

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Proceedings of the 7th International Sympo- sium* *on the Acquisition of Second Language Speech. Concordia University Working Papers in Applied Linguistics, 5*, 529–46.

Rato, A., Rauber, A. S., Kluge, D. C., & Santos, G. R. (2015). Designing speech perception tasks with TP. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 295–313). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Renard, R. (1979). *Introduction à la méthode verbo-tonale de correction phonétique. Troisième édition entièrement refondue*. Bruxelles, Belgium: Centre International de Phonétique Appliquée, Didier / Mons.

Reynolds, M., & Jefferson, L. (1999). Natural and synthetic speech comprehension: comparison of children from two age groups. *Augmentative and Alternative Communication, 15*(3), 174–182.

Rinaldo, S. B., Tapp, S., & Laverie, D. A. (2011). Learning by tweeting: Using Twitter as a pedagogical tool. *Journal of Marketing Education, 33*(2), 193–203.

Roach, P. (2000). *English phonetics and phonology*. Cambridge, UK: Cambridge University Press.

Rochdi, Y., & Mora, J. C. (2012). Mimicking foreign-accented speech as evidence of category formation for L2 sounds. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà & R. Mateu Serra (Eds.), *La lingüística aplicada en la era de la globalización*. Actas del XXX Congreso Internacional AESLA (pp. 96–101). Lleida, Spain: Universitat de Lleida.

Rochet, B. (1995). Perception and production of L2 speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 379–410). Timonium, MD: York Press.

Rosell-Aguillar, A. (2007) Top of the pods: In search of a podcasting "pedagogy" for language learning. *Computer Assisted Language Learning*, *20*(5), 471–492.

Rosell-Aguilar, F. (2009). Podcasting for language learning: Re-examining the potential. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 13–34). San Marcos, TX: CALICO.

Rubin, D. L. (1992). Non-language factors affecting undergraduates' judgements on non-native English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lectura topi con undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations, 14,* 337–353.

Ruellot, V. (2011). Computer-assisted pronunciation learning of French /u/ and /y/ at the intermediate level. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference* (pp. 199–213). Ames, IA: Iowa State University.

Saito, K. (2013). Re-examining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*, 1–29.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*, 595–633.

Salaberry, R. M. (1996). A theoretical foundation for the development of pedagogical tasks in computer mediated communication, *CALICO Journal*, *14*(1), 5–34.

Sánchez, A. (2009). La enseñanza de idiomas en los últimos cien años. Métodos y enfoques. Madrid, Spain: SGEL.

Saran, M., Seferoglu G., & Cagiltay K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research, 34*, 97–114.

Sato, C. J. (1991). Sociolinguistic variation and language attitudes in Hawaii. In J. Cheshire (Ed.), *English around the world. Sociolinguistic Perspectives* (pp. 647–663). Cambridge, UK: Cambridge University Press.

Schairer, K. (1992). Native speaker reaction to non-native speech. *Modern Language Journal, 76,* 309–319.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore.

Schumann, H. H., Holroyd, J., Campbell, R. N., & Ward, F. A. (1978). Improvement of foreign language pronunciation under hypnosis: A preliminary study. *Language Learning, 28,*143–148.

Schwab, E., Nusbaum, H., Pisoni, D. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, *27*, 395–408.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19,* 245–254.

Scovel, T. (1988). *A time to* speak. A psycholinguistic inquiry into the critical period for human speech. Cambridge, MA: Newbury House.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics, 20*, 213–223.

Seedhouse, P., Preston, A., Olivier, P., Jackson, D., Heslop, P., Balaam, M., Rafiev, A., & Kipling, M. (2014). The European digital kitchen project. *Bellaterra Journal of Teaching & Learning Language & Literature*, *7*(1), 1–16.

Seedhouse, P., Preston, A., Olivier, P., Jackson, D., Heslop, P., Plötz, T., Balaam, M., & Ali, S. (2013). The French digital kitchen: Implementing task-based language teaching beyond the classroom. *International Journal of Computer Assisted Language Learning and Teaching, 3*(1), 50–72.

Sekiyama, K., & Tohkura, Y. (1993). Interlanguage differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21,* 427–444.

Setter, J. (2008). Theories and approaches in English Pronunciation. In R. Monroy & A. Sánchez (Eds.), *25 Years of Applied Linguistics in Spain: Milestones and Challenges,* (pp. 447–457). Murcia, Spain: Universidad de Murcia, Servicio de Publicaciones.

Setter, J., & Jenkins, J. (2005). Pronunciation. *Language Teaching, 38,* 1–17.

Sharma, P., & Barret, B. (2007). Blended learning: Using technology in and beyond the language classroom. Oxford, UK: MacMillan Publishers Limited.

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165–180.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3,* 243–261.

Shockey, L. (2003). *Sound patterns of spoken English.* Malden, MA/Oxford, UK: Blackwell Publishing Ltd.

Siciliano, C., Faulkner, A., & Williams, G. (2003). Lipreadability of a synthetic talking face in normal hearing and hearing-impaired listeners. *Proceedings of the International Conference on Audio-Visual Speech Processing 2003*, 205–208.

Singleton, D. (1995). Introduction: A critical look at the critical period hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 1–29). Clevedon, UK: Multilingual Matters.

Singleton, D. (2003). Critical period or general age factor(s)? In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 3–22). Clevedon, UK: Multilingual Matters.

Skehan, P. (1989). Individual differences in second-language learning. London, UK: Edward Arnold.

Smith, J., & Beckmann, B. (2005). Improving pronunciation through noticing reformation tasks. In J. Maidment (Ed.), *Proceedings of the Phonetics Teaching and Learning Conference 2005.* (CD Rom edition). London, UK: University College London. Retrieved from: http://www.phon.ucl.ac.uk/home/johnm/ptlc2005/pdf/ptlcp46.pdf

Smith, J., & Beckmann, B. (2010). Noticing-reformulation tasks as a stimulus towards continuous autonomous phonological development. *New Zealand Studies in Applied Linguistics, 16*(1), 36–51.

Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal, 10*(3), 19–30.

Spada, N., & Lightbown, P. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly*, *42*, 181–207.

Stanley, G. (2006). Podcasting: Audio on the Internet comes of age. *TESL-EJ*, *9*(4), 1–7.

Stevens, K. N. (1985). Evidence for the role of acoustic boundaries in the perception of speech sounds. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in the honor of Peter Ladefoged* (pp. 243–255). Orlando, FL: Academic Press.

Stevenson, M. P., & Liu, M. (2010). Learning a language with web 2.0: Exploring the use of social networking features of foreign language learning websites. *CALICO Journal, 27*(2), 233–259.

Stockwell, G. (2007) Vocabulary on the move: investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning, 20*(4), 365–383.

Stockwell, G. (2010). Using mobile phones for vocabulary activities: examining the effect of the platform. *Language Learning & Technology, 14*(2), 95–110. http://llt.msu.edu/vol14num2/stockwell.pdf

Stockwell, G. (2012). Commentary: Working with constraints in mobile learning: A response to Ballance. *Language Learning & Technology, 16*(3), 24–31. http://llt.msu.edu/issues/october2012/stockwell.pdf

Stockwell, G. (2013). Tracking learner usage of mobile phones for language learning outside of the classroom. In P. Hubbard, M. Schultz, & B. Smith (Eds), *Human-computer interaction in language learning: Studies in honor of Robert Fischer* (pp. 118–136). San Marcos, TX: CALICO.

Stockwell, G., & Stockwell, E. S. (2012, August). Do smart phones affect learners' preferences for language learning? Paper presented at the EUROCALL 2012 Conference, Gothenburg, Sweden.

Stockwell, G., & Liu, Y-C. (2015). Engaging in mobile phone-based activities for learning vocabulary: An investigation in Japan and Taiwan. *CALICO Journal, 32*(2), 299–322.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*, 131–145.

Strik, H., Truong, K., de Wet, F., & Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication, 51,* 845–852.

Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning, 30*, 271–287.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–256). New York, NY: Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–484). Mahwah, NJ: Erlbaum.

Swan, K. (2003). Learning effectiveness: what the research tells us. In J. Bourne & J. C. Moore (Eds.) *Elements of quality online education, practice and direction* (pp. 13–45). Needham, MA: Sloan Center for Online Education.

Szyszka, M. (2011). Foreign language anxiety and self-perceived English pronunciation competence. *Studies in Second Language Learning and Teaching, 1*(2), 283–300.

Tan, E. S. Q., & Teo, D. J. L. (2015). Appsolutely smartphones: Usage and perception of apps for educational purposes. *Asian Journal of the Scholarship of Teaching and Learning, 5*(1), 55–75.

Taniguchi, M., & Abberton, E. (1999). Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language: work in progress*, *11*, 76–89. Retrieved from: https://www.phon.ucl.ac.uk/home/shl11/pdf_files/taniguchi.pdf

Tanner, M., & Landon, M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology, 13*(3), 51–65. Retrieved from: http://llt.msu.edu/vol13num3/tannerlandon.pdf

Taylor, P. (2009). *Text-to-speech synthesis*. Cambridge, UK: Cambridge University Press.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28,* 744–765.

Thornton, P., & Houser, C. (2001). Learning on the move: Vocabulary study via e-mail and mobile phone SMS. In C. Montgomerie & J. Viteli (Eds.), *Proceedings of EdMedia 2001* (pp. 1896–1897). Chesapeake, VA: Association for the Advancement of Computing in Education.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, *21*(3), 217–228.

Tomé, M. (2010). Enseñanza y aprendizaje de la pronunciación de una lengua extranjera en la web 2.0. *Revista de Lingüística y Lenguas Aplicadas, 5,* 221–239. Retrieved from: http://www.upv.es/dla_revista/docs/art2010/16_M_Tome.pdf

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16,* 183–203.

Trubetzkoy, N. (1958/1939). *Grundzuge der Phonologie* (Travaux du Cercle Linguistique de Prague 7). Gottingen, Germany: Vandenhoek & Ruprecht. (Original work published in 1939).

University of Iowa Research Foundation (UIRF). (2014). Sounds of Speech. (Version 1.6.5) [Mobile application software]. Retrieved from: https://itunes.apple.com/us/app/sounds-of-speech/id780656219?mt=8

Uther, M., Uther, J., Athanasopoulos, P., Singh, P., & Akahane-Yamada, R. (2007). Mobile adaptive CALL (MAC). A lightweight speech-based intervention for mobile language learners. *Proceedings of INTERSPEECH 2007*.

Vardanian, R. M. (1964). Teaching English through oscilloscope displays. *Language Learning, 14*(3-4), 109–118.

Walker, R. (2005). Using student-produced recordings with monolingual groups to provide effective individualized pronunciation practice. *TESOL Quarterly, 39*, 550–558.

Walker, N. R., Trofimovich, P., Cedergren, H., & Gatbonton, E. (2011). Using ASR technology in language training for specific purposes: A perspective from Quebec, Canada. *CALICO Journal, 28*(3), 539–552.

Wang, X., & Munro, M. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539–552.

Wang, S., & Smith, S. (2013). Reading and grammar learning through mobile phones. *Language Learning & Technology, 17*(3), 117–134. Retrieved from: http://llt.msu.edu/issues/october2013/wangsmith.pdf

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, 4 (2), 123–130.

Weinberg, A., Knoerr, H., & Vandergrift, L. (2011). Creating podcasts for academic listening in French: Student perceptions of enjoyment and usefulness. *CALICO Journal, 28*(3), 588–605.

Weinreich, U. (1953). *Languages in contact.* The Hague, The Netherlands: Mouton.

Wells, J. C. (2008). *Longman pronunciation dictionary*. Harlow, UK: Longman.

Wichern, P. U. M., & Boves, L. (1980). Visual feedback of Fo curves as an aid in learning intonation-contours. *Proceedings of the Institute of Phonetics Nijmegen 4*, 53–63.

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication*, *51*, 1024–1037.

Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.) *Handbook of research in second language acquisition* (pp. 671–691). Mahwah, NJ: Erlbaum.

Wilson, I. (2008). Using Praat and Moodle for teaching segmental and suprasegmental pronunciation. In T. Koyama (Ed.), *Proceedings of the WorldCALL 2008 Conference: CALL bridges the world.* (pp. 112–115). The Japan Association for Language Education and Technology.

Winters, S., & Pisoni, D. (2004). Some effects of feedback on the perception of point-light and full-face visual displays of speech: a preliminary report. In *Research on Spoken Language Processing Progress* Report No. 26, 139–164. Bloomington, IN: Speech Research Laboratory, Indiana University.

Witt, S. M. (1999). *Use of speech recognition in computer-assisted language learning.* PhD Thesis. Department of Engineering, University of Cambridge, Cambridge, UK.

Witt, S. M., & Young, S. (1997). Language learning based on non-native speech recognition. In *Eurospeech '97. Proceedings of the 5th European Conference on Speech Communication and Technology* (pp. 633–636). Rhodes, Greece, 22-25 September 1997.

Wong, J. W. S. (2015). Does proficiency matter? The effects of L2 vowel training on Cantonese ESL learners with high and low proficiency levels. In J. A. Mompean & J. Fouz-González (Eds.). *Investigating English pronunciation: Current trends and directions* (pp. 219–240). Basingstoke, UK/NewYork, NY: Palgrave Macmillan.

Wong, W. (2005) *Input enhancement: From theory and research to the classroom*. New York, NY: McGraw-Hill.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, *52*, 376–392.

Yule, G., Hoffman, P., & Damico, J. (1987). Paying attention to pronunciation: The role of self-monitoring in perception. *TESOL Quarterly, 21*, 765–768.

Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantisation. *Hispania, 77*(3), 470–481.

Zampini, M. L. (1996). Voiced stop spirantization in the ESL speech of native speakers of Spanish. *Applied Psycholinguistics, 17*, 335–354.

Zampini, M. L. (1997). L2 Spanish spirantization, prosodic domains and interlanguage rules. In S. J. Hannahs & M. Scholten (Eds.), *Focus on phonological acquisition* (pp. 209–234). Amsterdam, The Netherlands: John Benjamins.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design.* Sebastopol, CA: O'Reilly.

Zielinski, B. W. (2008). The listener: no longer the silent partner in reduced intelligibility. *System, 36*, 69–84.

Index