On the relationship between native and non-native speech perception and speech production

Judith Schmitz

TESI DOCTORAL UPF / 2015

SUPERVISORS

Dra. Núria Sebastián Gallés,
Dra. Begoña Díaz Menéndez

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

Universitat
Pompeu Fabra
Barcelona

ACKNOWLEDGMENTS

Thank you to Núria and Bego for giving me the chance to do the PhD project. Thank you for your support and encouragement, even when results did not look as expected. I learned a lot during my dissertation and I wouldn't have been able to accomplish it without you.

Thank you to Luciano, Ale, Laura, and Ele for giving me the chance to stay in your lab, to be able to perform the TMS study and the fun and new experiences. I will never forget the afternoon that you introduced me to the new experience of seeing flashes after applying TMS to the occipital lobe.

Thank you to everyone that helped me with my studies. Thanks to Karla, Robby, Mireia, and Aina for helping me to collect data. Without you I would still be testing participants! Thank you to Yu, Anna, and Alba for teaching me how to collect and analyze EEG data, to Miguel for introducing me to factor analysis and to Yu, Alba, Mathilde and Miguel for reading some of my drafts. Thank you to Carolina and Salva for sharing your stimuli and experimental program with me. Thank you to Xavi, Silvia, Cristina, and Florencia for the technical and administrative support.

Thank you to all the old and new members of the SAP group for your comments and suggestions in my presentations in the seminar and for listening to the same project over and over again!

Thank you to my friends at the UPF for sticking with me through the good and bad times of the PhD. Just remember: If I can finish, so can you!

Gracias a mi familia y amigos en Barcelona por sentirme aquí en casa.

Vielen Dank an meine Familie und Freunde in Deutschland. Vielen Dank für die vielen Besuche und Treffen. Ich freue mich jedesmal wieder wenn wir uns sehen und dass ihr es jedesmal wieder schafft, dass wir und treffen können – selbst wenn ich mal wieder viel zu spät Bescheid sage wann ich komme.

Thank you to Angel for teaching me how to make lasagna from spaghetti code, for your patience and your support throughout this time. But especially thank you for making me laugh when I was down and for always believing in me.

ABSTRACT

Models of speech perception differ in the nature of the relationship between speech perception and production. Whether speech perception and production processes are based on a common representations— the articulatory gesture— or speech perception fundamentally operates on the acoustic code is highly debated. In three experimental studies, we investigated the nature of the relationship between speech perception and production. In the first study we found an active role of the speech production system in speech perception, even when listening to unfamiliar phonemes. In the second study we found no influence of a somatosensory manipulation applied to an articulator in passive speech perception. In the third study we showed that speech perception and production abilities are tightly related across phonological processes (sub-lexical and lexical) and participants' languages (native— L1— and second language— L2—). The results suggest that speech perception and production are intimately linked.

RESUMEN

Los modelos de la percepción del habla difieren sobre la naturaleza de la relación entre la percepción y la producción del habla. El debate se centra en si ambos procesos comparten como representación básica los gestos articulatorios o bien si la percepción del habla se basa en el código auditivo. Investigamos la naturaleza de la relación entre la percepción y producción del habla en tres estudios experimentales. El primer estudio mostró que el sistema de producción del habla participa activamente en la percepción. El segundo estudio no reveló influencias en la percepción pasiva del habla de una manipulación somatosensorial aplicada en un articulador. El tercer estudio mostró una fuerte relación entre las habilidades de la percepción y producción del habla en varios procesos fonológicos (sub-léxicos y léxicos) y lenguas conocidas por los participantes (primera y segunda lenguas). Los resultados sugieren que la percepción y producción del habla están íntimamente relacionadas.

PREFACE

We are listening to speech constantly, when talking to a friend, calling someone on the phone, or listening to the radio. Understanding speech is an automatic and effortless process. We are able to listen to someone at the same time that we can do other things, such as walking or driving the car. However, when we consider the acoustic properties of the speech signal one can realize that analyzing speech and categorizing speech sounds is actually a very complex process. The acoustic properties of phonemes vary not only among speakers but also for the same speaker. We are all having different voices and this influences how the speech sounds. For instance, generally speaking, when a woman or a man say the same speech sound, let's say the vowel "a", it sounds much higher when a woman says it than when a man says it. This means that the acoustic signal for the "a" articulated by a woman and the "a" articulated by a man are acoustically very different. But even among women and men we find many acoustic differences in their speech sounds. Speech sounds are also acoustically different even for the same speaker because of co-articulation that causes influences of preceding speech sounds on the next ones, speech rate, or even emotional state.

The speech perception system faces the task of finding, among so much acoustic variability, those invariant properties that characterize each phoneme category. Despite that we can recognize speech sounds fast, robustly, and effortless, even now computer algorithms struggle to achieve human performance in categorizing speech sounds. So then ***how is our brain able to perceive speech*** so easily while it is difficult to decode for computers?

Historically, researchers proposed two different accounts for speech perception, an acoustic and a motor and one. As the name already says, acoustic models of speech perception focus on the acoustic nature of speech (Diehl, Lotto, & Holt, 2004; Flege, 1995; Hickok & Poeppel, 2007; Klatt, 1979; Kuhl & Iverson, 1995; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). Their main goal is to find algorithms that are able to interpret the acoustic waveform and extract from it smaller speech units, such as phonemes and words.

The motor account on the other hand focuses on similarities between speech perception and production (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010). Thus, the motor view proposes that speech perception and production share a common representation of speech sounds - the articulatory gesture. This means that when listening to speech, we translate the acoustic signal into an articulatory gesture to be able to process speech and to identify phonemes and words. Liberman and Mattingly (1985) proposed articulatory gestures as the basic unit for speech perception because articulatory gestures vary less than the acoustic signal, meaning that the speech perception system can easily find the invariant properties of speech sounds that are characteristic of the phoneme categories.

However, this assumption has been challenged by acoustic accounts that claim that speech perception is an auditory computation and that taking into account a combination of acoustic cues can lead to the identification of speech sounds without translating them into the articulatory gesture (Diehl et al., 2004).

Recent neuroimaging studies opened up again the debate about whether speech perception relays on acoustic or motor processing by showing influences of speech production on speech perception (D'Ausilio et al., 2009; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Möttönen & Watkins, 2009; Pulvermüller et al., 2006; Watkins, Strafella, & Paus, 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004). The main goal of this dissertation is to test the ***relationship between speech perception and production*** proposed by the motor accounts of speech perception.

This dissertation is divided into three sections. In the first section (Chapter 1: "*Introduction*"), we review in detail the different theories and models for the motor and acoustic accounts of speech perception. We also review three recent neurobiological experimental approaches that challenge the acoustic view of speech perception by claiming that brain areas involved in language production are an integral part of the speech perception system. The first set of experimental series showed that listening to speech does not only activate brain areas related to speech perception, but also brain areas related to speech production (Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003; Wilson et al., 2004).

These studies investigated patterns of neural activations when perceiving speech and we will group them in the section "*The influence of speech perception on speech production*". A second set of studies reported that manipulating the production system influenced speech perception (D'Ausilio et al., 2009; Ito, Tiede, & Ostry, 2009; Möttönen & Watkins, 2009; Sato et al., 2011). These studies focused on the effects of speech production in speech perception and thereby they will be presented under "*The influence of speech production on speech perception*". Finally, we review a third set of studies that measured the strength of the relationship between speech perception and production by correlating the two processes directly (Flege, MacKay, & Meador, 1999; Kartushina & Frauenfelder, 2014; Levy & Law, 2010; Peperkamp & Bouchon, 2011; Rallo Fabra & Romero, 2012). These studies were usually performed with second language (L2) learners because they show higher variability in their perception and production abilities compared to native speakers who perform at ceiling. The high variability in L2 learners' abilities allows for the use of correlations, which tests the relationship between speech perception and production in both directions. These studies are described in the section "*Correlations between speech perception and production skills*". We end this chapter by providing an overview of the studies conducted within this dissertation and how they extend the previous literature.

The second section includes three experimental studies, submitted to international peer-reviewed journals. In the first study (Chapter 2:

“*Increased lip corticobulbar excitability during the perception of non-native phonemes*”) we explore if listening to native and non-native phonemes activates the speech production system by applying transcranial magnetic stimulation (TMS) over the lip and tongue representations in the motor cortex and recording excitability of the corticobulbar tract (connecting the motor cortex with facial muscles). The second study (Chapter 3: “*Attention modulates somatosensory influences in passive speech listening*”) investigates if constraining the production system by means of a somatosensory manipulation, a spoon over the tongue, influences speech perception of native and non-native phonemes in passive listening conditions by measuring electrophysiological responses (EEG). In the third study (Chapter 4: “*Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities*”) we test the strength of the relationship between speech perception and production by measuring speech perception and production processes across phonological processes (sub-lexical and lexical), languages (native, L2, and foreign), and sensory modalities (auditory and visual).

In the last section (Chapter 5: “*Discussion*”) the results of the three experimental studies are combined to generate a general discussion. We first summarize the three studies and how their results contribute to the current literature. Then we report the general conclusions based on all three studies and propose further lines of research.

TABLE OF CONTENTS

1. INTRODUCTION

Speech perception and production relationship

What is the nature of the relationship between speech perception and production is still highly debated and the basic code of the speech perception representations remains elusive. Traditionally, researchers in the field of language proposed that speech perception and production were two different processes and, consequently, they studied them separately. Early studies based this conclusion on the observation that patients with brain damage in the frontal lobes had problems in speech production (Broca's aphasia), while patients with brain damage in the temporal lobe showed problems in speech perception (Wernike's aphasia) (see Gazzaniga, Ivry, and Mangun (2002) for a review). This assumption led to the development of separate models for speech perception (Klatt, 1979; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994) and production (Dell, 1986; Fromkin, 1971; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Yet, other views conceptualized speech perception and production as closely related processes that share a common articulatory representation of phonemes (Fowler, 1986; Liberman & Mattingly, 1985).

One of the first theories that claimed for a common representation of speech perception and production processes was the Motor Theory of Speech Perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The theory was based on the observation that the acoustic signal for the same phoneme can vary depending on the phonemes that surround it in a word. Thus, the lack of invariant features for speech sounds in the auditory speech signal led researchers to wonder how the speech perception system dealt with such a high variability. Liberman et al. (1967) reasoned that the basic unit of speech perception should be an invariant element and proposed it to be the articulatory gesture. According to the Motor Theory, speech perception was thus explained by abstracting the articulatory gesture from the speech signal, providing the connection between speech perception and production mechanisms. Some researchers challenged such assumption by claiming that the listener could also learn to associate different acoustic cues for the same phoneme category and integrate them (Diehl et al., 2004).

The motor view of speech perception gained again attention by recent neuroimaging studies showing that listening to speech activates brain networks extending beyond the traditional auditory areas within the temporal lobes, and included speech production areas, such as motor and somatosensory cortices (Pulvermüller et al., 2006; Wilson et al., 2004). Based on this evidence, researchers proposed more recent neurobiological models of speech processing (see Figure 1.1 for an example) which allow for interactions between speech perception and production areas in the brain (Hickok & Poeppel, 2007). In this model, speech production is not essential for perception, but rather would be recruited only when the listening conditions are difficult.

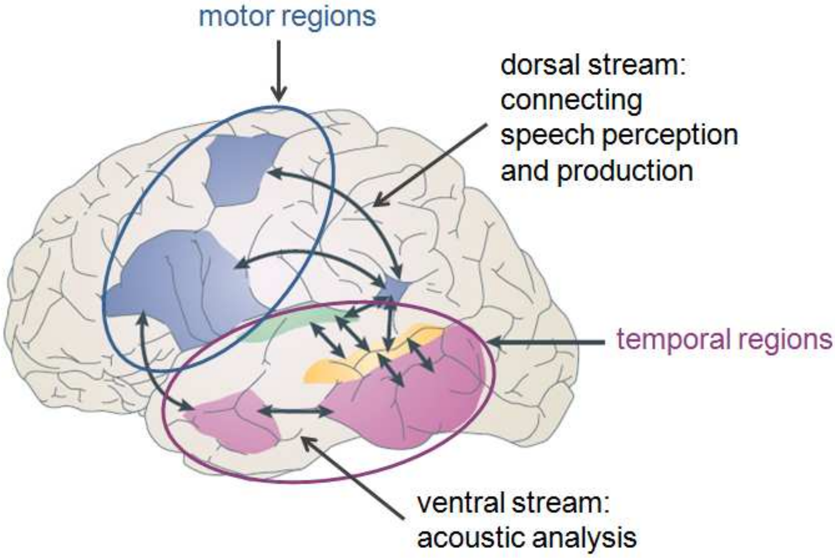

Figure 1.1. *Representation of the brain areas involved in speech perception in the left hemisphere according to the Dual Stream Model (Hickok & Poeppel, 2007). Inside the purple oval are the traditional speech perception areas in the temporal lobe, connected through the ventral stream which is proposed to analyze the acoustic signal. Inside the blue oval are the traditional speech production areas in the motor regions of the brain. Speech perception and production areas are suggested to connect through the dorsal stream. Reproduced from Hickok & Poeppel, (2007).*

Despite this recent evidence for the involvement of speech production areas in speech perception, the nature of the relationship between speech perception and production remains still elusive (Hickok, Houde, & Rong, 2011; Pulvermüller & Fadiga, 2010; Scott, McGettigan, & Eisner, 2009). In this dissertation we contribute to the ongoing debate by investigating the relationship between the two linguistic processes under varying degrees of language proficiency, from the native language, to a second language (L2), to a completely unknown language (hereafter referred to as “non-native”). In this chapter we review different theories and models of speech perception according to their conceptualization of the nature of the basic speech unit (motor vs. acoustic) and provide different experimental approaches that have been previously employed to tackle the nature of the relationship between speech perception and production.

1.1. Theories and models of speech perception

In general, two broad categories of speech perception theories and models can be distinguished: motor and acoustic ones, each roughly corresponding to contrasting views about the nature of the relationship between perception and production. While motor accounts assume that speech perception and production share a common representation of speech sounds: the articulatory gesture (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010), acoustic accounts assume separate representations for speech perception and production, connecting at some point (often an implicit assumption) (Diehl et al., 2004; Flege, 1995; Hickok & Poeppel, 2007; Klatt, 1979; Kuhl & Iverson, 1995; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). This means that according to the motor view, not only when producing speech, but also when listening to speech, the articulatory gesture is crucial for the identification of the perceived speech sounds. Hence, the reliance on common representations leads to a tight link between speech perception and production and implies a mutual dependency between speech perception and production. In contrast, according to the acoustic view of speech perception, the acoustic signal contains enough information to perceive speech and speech production does not play a role in perception. Some acoustic models argue that being able to perceive a sound accurately is a necessary requirement for being able to produce it (Flege, 1995; Kuhl & Iverson, 1995) and, hence, posit a dependence of production on perception, at least when learning a language. The acoustic view of speech perception, thus, expects only an influence of speech perception on speech production.

1.1.1. Motor theories and models

The two most representative motor theories are the revised Motor Theory of Speech Perception by Liberman and Mattingly (1985) and the Direct Realist Theory by Fowler (1986). According to the revised Motor Theory of Speech Perception, the listener perceives the intended articulatory gesture of the speaker. This process is performed by a language-specific and innately given representation of the vocal tract synthesizer that contains information about its anatomy and the mapping between acoustic signals and articulatory speech gestures. When listening to speech, the synthesizer translates the acoustic signal into the articulatory gesture that is further decoded into neuromotor commands corresponding to specific features of movements (e.g. “labial”, “occlusive”, “nasal”, etc.). Phonemes are then identified by a specific combination of neuromotor commands.

In the Direct Realist Theory, Fowler (1986) argues that the articulatory gestures are directly obtained from the acoustic signal and no transformation of the gestures, such as neuromotor commands in the Motor Theory, has to take place. The Direct Realist Theory proposed that each articulatory gesture structures the acoustic signal in a specific manner thus providing enough information to the listener to recover the articulatory gesture directly from the auditory input. Through general learning mechanisms, the listener attunes to the information that is relevant to abstract articulatory gestures from the acoustic signal and actively searches for this information in it. Thus, the Direct Realist Theory is based on general perception and learning mechanisms, in contrast to the Motor Theory postulation that a language specific and innately given vocal tract synthesizer is needed for speech perception.

Best (Best, 1995; Best & Tyler, 2007) implemented the Direct Realist Theory for non-native speech perception in the Perceptual Assimilation Model (PAM). According to PAM, non-native phonemes are perceived by comparing them to the articulatory gestures of native phonemes and finding similarities and/or discrepancies between them. Based on the similarities, the phonemes can be assimilated to a native phoneme category or be assimilated either as an uncategorizable speech or as a non-speech category. If two non-native phonemes are assimilated to the same native category, discrimination is expected to be difficult. However, if two non-native phonemes are assimilated to different categories, or they differ in how good they fit within the same category, they should be properly perceived. These predictions apply to non-native perception and production skills, as both processes rely on the articulatory gesture.

A sub-set of motor models and theories claims that the representations of speech are based on articulatory information but they also acknowledge that sensory input is relevant for speech perception. This multimodal view of speech perception is represented by the Perception-for-Action-Control Theory (PACT) (Schwartz, Basirat, Ménard, & Sato, 2012) and the Fuzzy Logical Model of Perception (FLMP) (Massaro, 1987). In PACT, Schwartz et al. (2012) claim that the basic units of speech perception convey multimodal cues such as articulatory, auditory, and visual information. Based on the observation that both speech perception and production shape the speech sounds, the model assumes that we perceive “perceptually-shaped gestures” or “perceptuo-motor units”. The FLMP by Massaro (1987) claims that speech units rely on both acoustic and visual speech information. According to this model, phonemes are represented in the form of prototypes, which represent the perfect auditory (e.g., acoustic information such as formant frequencies) and visual values (e.g., visually seen articulatory gesture such as lip closing) for the phonemes. Therefore, FLMP links speech perception and production through the auditory and visual modalities. Hence, these multimodal accounts of speech perception expect a close relationship between speech perception and production but differ from the other motor theories and models in this section in regard to the basic code of speech representations.

Motor theories, such as the Motor Theory of Speech Perception (Liberman & Mattingly, 1985), are very vague about the mechanisms connecting speech perception and production. The discovery of mirror neurons postulated a possible solution for a neural mechanism integrating speech perception and production. These neurons were first discovered in the monkey. They are active during both the perception and execution of a movement and thus combine perception and production in the same entity (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). Furthermore, a set of audio-visual mirror neurons has been found to respond to the performance of an action, the perception of the action, and the sound produced by the action alone (Kohler et al., 2002). These neurons are found in a brain area of the monkey (area F5) that has been proposed to be the homologue of Broca's area in humans, an important brain region for speech production (Rizzolatti & Arbib, 1998). The finding that Broca's area had functional similarities to mirror neurons, such as an involvement in action perception and production, lead to the claim that humans possess a mirror system that integrates speech perception and production (Pulvermüller & Fadiga, 2010).

To sum up, the motor view of speech perception suggests that speech perception and production share a common representation: the speech articulatory gesture. The perceived acoustic input is translated into the articulatory gesture either by an innately given and language-specific vocal tract synthesizer (Liberman & Mattingly, 1985) or directly obtained from the acoustic signal (Best, 1995; Best & Tyler, 2007; Fowler, 1986). Others conceptualize speech representations as multimodal units that combine production and perception information (Massaro, 1987; Schwartz et al., 2012). The discovery of mirror neurons led to the proposal that these neurons might integrate speech perception and production in the same entity (Pulvermüller & Fadiga, 2010).

1.1.2. Acoustic theories and models

Acoustic models of speech perception propose that speech perception is only an acoustic process. The Lexical Access From Spectra (LAFS) by Klatt (1979) suggests that words are perceived directly from the acoustic signal–the spectrum. Other models including the Cohort Model by Marslen-Wilson and Welsh, (1978), the TRACE model by McClelland and Elman (1986), and the Shortlist Model by Norris (1994) propose that the acoustic signal is transformed into abstract phonemic representations which are then used to find words in the mental lexicon. These models focus solely on speech perception and do not address how it relates to production. However, this does not exclude the possibility that speech perception and production processes interact at some point during the speech perception process.

Examples of acoustic models that focus on phoneme learning are the Native Language Magnet Model from Kuhl and Iverson (1995) and the Speech Learning Model (SLM) from (Flege, 1995). The idea of the Native Language Magnet Model is that the acoustic space corresponding to different instances of the same phonetic category are dragged together and integrated to one single category. The central sound in the category is called the prototype and surrounding instances are perceived as closer to the prototype than two sounds with the same acoustic distance outside the prototype. Kuhl and Iverson claim that the prototypes are shaped by experience in early language acquisition and that the prototypes influence both speech perception and speech production.

An acoustic model of L2 phoneme learning is the Speech Learning Model (SLM) from Flege (1995). Similar to PAM (Best, 1995; Best & Tyler, 2007), SLM proposes that native and L2 speech sounds share a single phonetic space and that the learning of the L2 phonemes will depend on how they relate to the native ones. However, for SLM, L2 phonemes are related acoustically to native phonemes and not articulatory, as proposed by PAM. This difference in reliance on the acoustic or articulatory space leads to different predictions about the relationship between L2 perception and production skills. As described before, in PAM, Best proposes that L2 perception and production are both depended on the articulatory gesture. This means that accurate L2 phoneme perception and production abilities are simultaneously developed and both abilities should be highly correlated. In contrast, Flege suggests in SLM, that accurate L2 phoneme perception precedes accurate L2 production. This means that, especially in early learning phases, L2 learners may perceive a phoneme accurately but not produce it, leading to imperfect correlations. In later learning phases L2 production skills are expected to improve, leading to moderate to high correlations between L2 perception and production skills.

More recent acoustic theories and models of speech perception, such as the Dual Stream Model for speech perception from Hickok and Poeppel (2007), postulate that speech perception and production processes can interact under some circumstances. In the Dual Stream Model (see Figure 1.1), Hickok and Poeppel suggest a direct functional connection between speech perception and production brain areas. The Dual Stream model consists of two streams: a ventral and a dorsal one. The ventral stream, involving structures of the temporal lobe, analyzes the acoustic signal and retrieves the linguistic meaning by breaking the acoustic signal into its components of phonemes, words, etc. A second stream, the dorsal stream, links auditory perception areas in the temporal lobe with articulatory motor areas in the frontal regions of the brain through the parietal areas. The dorsal stream thus allows for interactions between perception and production abilities. Yet, the authors argue that the dorsal stream is only activated when the auditory signal alone is not sufficient for an accurate perception (Hickok et al., 2011).

An acoustic approach that directly argues against motor theories is the General Auditory and Learning Approach by Diehl et al. (2004).

Diehl and colleagues propose that speech perception and production processes are related because the auditory signal and articulatory gestures are repeatedly presented together, leading to correlations between them. However participants can also be taught to identify phonemes by repeatedly seeing correlations between meaningless auditory and tactile information. Since in this case phonemes can be identified by other information than articulatory gestures, Diehl and colleagues claim that correlations between the acoustic signal and articulatory gestures do not proof that speech perception is articulatory.

To sum up, according to acoustic theories and models, speech perception relies exclusively on the acoustic analysis of the speech signal (Diehl et al., 2004; Flege, 1995; Hickok & Poeppel, 2007; Klatt, 1979; Kuhl & Iverson, 1995; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). Some acoustic models of language learning postulate a link between the two abilities and claim that speech production depends on the phoneme categories created by speech perception (Flege, 1995; Kuhl & Iverson, 1995). Furthermore, more recent models allow for influences between speech perception and production processes under some circumstances, such as difficult listening conditions (Hickok & Poeppel, 2007). Other researchers, such as Diehl and colleagues (2004) propose that correlations between the acoustic signal and articulatory gestures are established but that these correlations do not proof that articulatory gestures are the basic unit of speech perception. Thus, in contrast to motor theories, for which speech production is essential for speech perception, acoustic theories suggest that speech production depends on perception.

In the next section we describe three lines of research that aimed at investigating the nature of the relationship between speech perception and production.

1.2. Experimental evidence for the interaction between speech perception and production

In this section we describe three experimental approaches that tested the predictions of the motor and acoustic accounts on the nature of the relationship between speech perception and production. According to the motor view, speech perception and production share the same set of phoneme representation (i.e., gestures) and, hence, are intimately linked (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010). Empirically, this claim implies bidirectional influences between speech perception and production, that is, speech perception should influence production but also speech production should influence speech perception. In contrast, acoustic accounts of speech perception (Diehl et al., 2004; Flege, 1995; Hickok & Poeppel, 2007; Klatt, 1979; Kuhl & Iverson, 1995; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994) claimed that speech perception is an auditory process and does not require production. According to the acoustic account, speech perception can influence speech production because production depends on the phoneme representations build up by the speech perception system (Flege, 1995; Kuhl & Iverson, 1995) but only under difficult listening situations, when the perception system is taxed, the production system intervenes in perception as a compensatory mechanism (Hickok & Poeppel, 2007).

The first line of research studies reviewed in this section investigated the effects of speech perception manipulations on the activity of production brain areas. The second set of studies tested the reverse manipulation, that is, if manipulating the activity of the motor brain areas had an impact on speech perception. Those studies were carried out in the native language of the participants. The third approach compared speech perception and production abilities in L2 learners to investigate if a correlation existed between the two skills. As explained in the previous paragraph, both motor and acoustic views agree in the predictions that speech perception manipulations may have an effect on production (first line of studies here reviewed) and that perception and production abilities correlate (third line of studies reviewed). Yet, that speech production manipulations may affect speech perception (second set of studies reviewed) is only predicted by the motor view of speech perception and, hence, is the critical evidence to refute the acoustic view. Note, however, that an acoustic model of speech perception, the Dual Stream Model (Hickok & Poeppel, 2007), stated that speech production may be necessary for perception only to overcome difficult listening conditions.

1.2.1. The influence of speech perception on speech production

Researchers investigated the influence of speech perception on production by measuring if listening to speech activates production areas in the brain. Some of these studies used functional Magnetic Resonance Imaging (fMRI) to show that listening to speech activated brain areas extending beyond traditional auditory areas to speech production areas in motor cortices (Pulvermüller et al., 2006; Wilson et al., 2004). For example, Pulvermüller et al. (2006) compared the brain activations when participants passively listened to or articulated the consonants ‘p’ or ‘t’, produced by a lip or a tongue movement respectively. Results showed that passively listening to the consonants resulted in activations in the same regions of the motor cortex that were also active when articulating the same consonants.

Similar results have also been obtained measuring corticobulbar excitability (Fadiga et al., 2002; Sundara, Namasivayam, & Chen, 2001; Watkins et al., 2003). To measure corticobulbar excitability, Transcranial Magnetic Stimulation (TMS) is applied over the motor cortex and the excitability of the corticobulbar tract (connecting the primary motor area with the facial muscles) is recorded by means of electrodes placed on the muscles. Watkins et al. (2003) recorded corticobulbar excitability from the lip muscle, while participants were viewing or listening to speech in comparison to two non-speech conditions: listening to non-verbal sounds (such as car engines) and viewing eye movements. The authors found enhanced corticobulbar excitability for the speech conditions compared to viewing eye movements. Listening to non-verbal sounds resulted in an intermediate excitability that was not different from the speech conditions, or from viewing eye movements (probably because some of the non-verbal sounds could be imitated). Murakami, Restle, and Ziemann (2011) showed that the corticobulbar excitability even increased during difficult speech viewing and listening tasks, such as when the speech was speeded up or embedded in noise. Furthermore, corticobulbar excitability studies showed that activations of the production areas during listening to speech can be specific to the articulatory muscles used during the production of the speech sounds. Fadiga, Craighero, Buccino, and Rizzolatti (2002) found that the corticobulbar excitability of the tongue muscle was higher during listening to words articulated with a tongue movement (‘rr’) compared to words that do not use a tongue movement (‘ff’).

The studies described before showed that listening to native speech activated not only speech perception but also production areas. However, it is less clear if listening to non-native speech can activate production areas, since both the acoustic and articulatory representations of non-native phonemes are poorly specified to the speaker. Evidence for activations of production areas during listening to non-native speech comes from an fMRI study. Wilson and Iacoboni (2006) found higher activations in production areas when participants listened to non-native phonemes, compared to native phonemes. In contrast to the higher activations for non-native speech found by the fMRI study, measuring cortiobulbar excitability, Swaminathan et al. (2013) found higher activations for watching visual native speech compared to non-native speech. In the study, lip corticobulbar excitability was measured while natives and L2 learners of English were viewing known speech (English), unknown speech (Hebrew), non-speech movements (gurns), and a static face. The results showed increased corticobulbar excitability for viewing known speech compared to viewing unknown speech or non-speech lip movements for both native and L2 learners. Surprisingly, viewing a static face also resulted in higher lip corticobulbar excitability that was not different from viewing known speech what was interpreted as an anticipation of hearing speech.

To sum up, as expected by the motor and acoustic accounts of speech perception, previous studies showed evidence for activations of the production system during native speech perception (Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003; Wilson et al., 2004), while it is not clear if these activations are dependent on sensory-motor experience. Although an fMRI study showed higher activations of the production areas when listening to non-native speech (Wilson & Iacoboni, 2006), another study measuring corticobulbar excitability showed higher activations for viewing known as compared to unknown speech (Swaminathan et al., 2013). Furthermore, it is not known if listening to non-native speech can result in articulatory specific activations in the motor cortex and if these activations might be related to other measures, such as individual perception and/or production skills.

1.2.2. The influence of speech production on speech perception

Several studies investigated whether manipulating speech production areas in the brain influenced speech perception (D’Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Sato et al., 2011). Investigating this directionality, i.e., an influence of speech production on perception, is very important because, as explained before, only the motor account on speech perception expects such an influence. These studies tested if the production system influenced speech perception by means of activating or deactivating areas in the primary motor cortex with TMS (D’Ausilio et al., 2009; Möttönen & Watkins, 2009) or manipulating an articulator directly with a somatosensory manipulation (Ito et al., 2009; Sato et al., 2011). For example, Sato et al. (2011) trained participants to either press their tongue against the anterior palate or protrude their lips before listening to phonemes articulated with the tongue (/t/) or with the lips (/p/) masked with noise. The motor training influenced the response bias (beta) towards responding more often to the phoneme that was articulated with the muscle that they trained before.

Although these studies showed that the production system influenced speech perception, finding an influence depended on paying attention to the stimuli (Möttönen, Dutton, & Watkins, 2013;

Möttönen, van de Ven, & Watkins, 2014), and the difficulty of the listening conditions and/or tasks (Alho et al., 2012; D'Ausilio, Bufalari, Salmas, & Fadiga, 2012). This is in line with acoustic proposals suggesting that the production system is recruited under certain conditions, such as difficult tasks and/or listening conditions that require attention (Hickok et al., 2011; Scott et al., 2009). Regarding attention, Möttönen et al. (2013, 2014) deactivated the lip representation in the motor cortex with repetitive (r)TMS and then measured the auditory event related potentials (ERPs) elicited by phonemes articulated with the lips or tongue during attentive and passive listening conditions. The authors found that when attention was paid to the stimuli, the deactivation of the lip representation resulted in articulatory specific effects, influencing only the auditory ERPs evoked by the lip articulated phonemes. However, during passive listening, the deactivation of the lip representation resulted in general effects, influencing the ERPs to both lip and tongue articulated phonemes. The authors suggested that attention modulated the specificity of the influences of the production system on speech perception.

Researches also showed that the difficulty of the listening conditions and/or task modulated whether the production system influences speech perception (Alho et al., 2012; D'Ausilio et al., 2012). For example, D'Ausilio et al. (2009, 2012) manipulated the difficulty of the listening conditions while stimulating with TMS either the lip or tongue representation in the motor cortex when participants listened to phonemes articulated with the lips or tongue.

In a difficult condition the audio was masked with noise whereas in an easy condition the audio was not manipulated. The authors found that only in the difficult condition, when the audio was masked, the TMS stimulation influenced the behavioral performance of the participants: Participants were faster and more accurate in responding which phoneme they had listened to.

In support to the motor view of speech perception, studies manipulating the speech production system found that speech production influenced perception. However, the findings could also be accounted by the acoustic view of speech perception as a compensatory mechanism for difficult listening conditions. Hence, the nature of the influence of speech production on perception remains poorly understood.

1.2.3. Correlations between speech perception and production skills

The relationship between speech perception and production has also been investigating by analyzing whether the two abilities correlate in L2 learners. In the native language it is very difficult to assess the correlation between speech perception and production skills since natives perform at ceiling and show little variability in these skills. On the contrary, L2 learners show large variability in their performance. Studies have shown differences in the strength of the relationship between L2 perception and production, ranging from high (Flege et al., 1999) to moderate (Levy & Law, 2010; Rallo

Fabra & Romero, 2012) to no correlations (Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011).

Flege et al. (1999) found high to moderate correlations between speech perception and production testing both abilities involving acoustic-phonological processing. L2 perception was measured in a discrimination task of isolated vowels and L2 production was accessed by repeating words spoken by a native speaker. Repetition tasks can be achieved by echoing the speaker and do not require access to any phonological representation of the stimuli. In contrast, the study by Kartushina and Frauenfelder (2014) showed no relationship between L2 perception and production when each ability was measured with tasks tapping onto different phonological processes. For L2 speech perception, an identification task of isolated vowels was used and, hence, low level acoustic-phonetic processing was assessed. L2 speech production instead was measured in two tasks, one involving acoustic-phonetic processing and the other lexical processing. In the acoustic-phonetic task, participants repeated vowels spoken by a native speaker. In the lexical task, they named the vowel present in a picture that required the retrieval of the lexical representation prior to the selection of the phonemes. Hence, the phonological processes accessed by the task to measure L2 perception and production were quite different.

The type of phonological processes compared is relevant because previous studies have shown that accurate phonological processes of L2 isolated vowels do not necessary result in accurate recognition of the L2 phonemes within words (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Sebastián-Gallés & Baus, 2005). The distinct mastery of phonological processes found by previous studies is in line with speech perception and production models, which propose a hierarchical structure of phonological processes (Dell, 1986; Fromkin, 1971; Hickok & Poeppel, 2007; Levelt et al., 1999; Levelt, 1989; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). For instance, in the case of speech perception, many sub-lexical processes are performed before accessing to the lexicon. The listener initially has to compute different acoustic-phonetic (spectral and/or duration information) and allophonic cues that are relevant to its own language to identify the isolated phonemes. The listener then can use phonotactic (i.e., rules about legal combination of phonemes within syllables and words) and stress cues that are relevant to his/her own language to determine what groups of isolated phonemes are a single lexical unit (i.e., a word). Then, the output of all this sub-lexical processes leads to the selection of the words in the mental lexicon.

The hierarchical structure of sub-lexical and lexical phonological processes implies that for speech perception accurate sub-lexical representations are needed for successful access to the mental lexicon and for speech production lexical selection is needed to select phonemes at the sub-lexical level. Hence, difficulties at one level of the hierarchy may percolate to the other levels. This prediction can easily be tested in L2 learners who show difficulties to discriminate between L2 phonemes. The prediction that difficulties at the phonological processing level hinders the ability to build and access words in the mental lexicon was supported by experimental data showing that good performance in sub-lexical tasks was a necessary requirement for early and late L2 learners to achieve good performance in tasks measuring lexical processes (Díaz et al., 2012; Sebastián-Gallés & Baus, 2005).

Sebastián-Gallés and Baus (2005) tested 80 early Spanish (L1)-Catalan (L2) bilinguals' difficulties to discriminate the Catalan /e/-/ɛ/ contrast in three tasks involving different phonological processes of the speech perception hierarchy. There is abundant evidence showing that native Spanish listeners have great difficulties in discriminating between the Catalan /e/-/ɛ/ vowel contrast, even if they started to learn Catalan at an early age. The difficulty would stem from the fact that the mid-front Spanish vowel /e/ falls roughly in between the two mid-front Catalan vowels /e/-/ɛ/ (Bosch, Costa, & Sebastián-Gallés, 2000; Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999). Sebastián-Gallés and Baus (2005) tested participants' sub-lexical processing by means of an identification task and an adaptation of the gating task. The categorization task involved the recognition of isolated vowels from a seven-step synthesized continuum. The gating task measured the identification of the Catalan /e/-/ɛ/ contrast in naturally uttered minimal word pairs, which includes the difficulty of co-articulation. The third task tested phonological lexical processes by means of an auditory lexical decision task. In this case, accurate acoustic-phonological processes were needed for accessing the words in the mental lexicon. Bilinguals' performance was compared to native Catalan listeners and showed a pattern that was congruent with the hierarchical structure of speech perception models: participants performing well in the lexical decision task also performed well in the sub-lexical tasks. However, only few bilinguals succeeded in the lexical decision task (18.3%) in contrast to higher percentage in the gating and categorization task (46.6% and 68.3% respectively) revealing that, surprisingly, accurate perception of isolated phonemes did not entail incorporating the accurate phonemic information in the lexicon.

These results were subsequently replicated by Díaz et al. (2012). In three tasks, the authors tested 55 late Dutch (L1)-English (L2) bilinguals' difficulties to discriminate between the English /æ/-/ɛ/ contrast, which has been shown to be difficult for Dutch natives (Cutler, Weber, Smits, & Cooper, 2004; Weber & Cutler, 2004). From the three tasks employed, two were the same as in Sebastián-Gallés and Baus (2005): a categorization task and a lexical decision task (Broersma & Cutler, 2011), but with English stimuli. The third task was a word identification task also measuring lexical phonological processes. In the word identification task, words were auditory presented at the end of a carrier sentence. The words belonged to minimal word pairs that differed only in the English /æ/-/ɛ/ contrast (such as "cattle" and "kettle"). At the same time than the auditory stimulus, pictures of both alternatives of the minimal word pair were presented and participants had to select the picture corresponding to the auditory word. Again, results were in line with the hierarchical structure, participants who performed well in the lexical task also performed well in the sub-lexical task but performing well in the sub-lexical task (43.63%) was not enough to succeed in the lexical tasks (lexical decision task: 12.72%, word identification task: 9.09%). Hence, measuring phonological ability at different levels of the hierarchy may be the cause for inconsistent findings in the correlation strength between L2 speech perception and production skills.

Although most studies investigated the correlation between speech perception and production in L2 learners, the results are meant to reveal the interaction between the two abilities regardless of the language familiarity of the participants. For instance, Flege et al. (1999) found a significant correlation between L1 speech perception and L2 speech production but the correlation strength was lower than for L2 speech perception and production. Perhaps, the little variability in L1 abilities caused the low correlation strength. Electrophysiological responses have shown to be more sensitive to capture variability than behavioral measures for L1 perception skills (Díaz, Baus, Escera, Costa, & Sebastián-Gallés, 2008) and may be the appropriate tool to study the correlation between L1 speech perception and production. Díaz et al. (2008) measured the electrophysiological Mismatch Negativity (MMN) response which is elicited when a change between a frequently repeated sound and an infrequent one is detected (Duncan et al., 2009; Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). The MMN amplitude has been related to discrimination ability between two sounds and has been proposed to be a measure of discrimination sensitivity (Amenedo & Escera, 2000).

Díaz et al. (2008) measured discrimination sensitivity to phonemes (detection of changes in native and unknown phonemes) and acoustic abilities (detection of changes in frequency, duration, and pattern) and tested if these abilities relate to L2 ability. L2 perception skills were measured in early Spanish (L1)-Catalan (L2) bilinguals using sub-lexical and lexical tasks involving the discrimination of the Catalan /e/-/ɛ/ contrast (a subset of participants in Sebastián-Gallés and Baus, 2005). Two groups of participants were selected to differ maximally in their L2 ability in both sub-lexical and lexical tasks. One group of “good” perceivers performed as well as Catalan natives in all L2 perception tasks while the other group of “poor” perceivers performed below Catalan natives in all tasks. The authors found that “good” and “poor” perceivers showed differences in the MMN between phoneme changes, regardless of participants’ familiarity with the phonemes. No differences between the groups were found for acoustic abilities. Díaz, Mitterer, Broersma, Escera, & Sebastián-Gallés (in press) reported parallel results with late Dutch (L1)-English (L2) bilinguals (AoA=12 years of age). The authors interpreted these findings as evidence for a speech-specific origin of L2 perception skills. The variability in phoneme discrimination captured by the MMN makes it a useful tool for testing if the perception and production relationship in the

L2 is influences by individual variability in perceiving native and unfamiliar phonemes.

Previous studies on the correlations between speech perception and production have been limited to the perception of acoustic stimuli. Yet, speech perception is a multisensory phenomena that entails the processing of auditory as well as visual information (MacDonald & McGurk, 1978; Massaro, 1987; Schwartz et al., 2012). Visual speech gestures are classified into visemes and one viseme can correspond to several phonemes that are articulated with the same (visible) articulation movement, such as the lip articulated phonemes /p, b, m/ (MacDonald & McGurk, 1978). Visemes have been shown to interact with phonemes in speech perception, such as in the McGurk effect (McGurk & MacDonald, 1976). In this effect, for instance, visually presented /ka/ interferes with mismatching auditory presented /pa/ and most participants report to have heard /da/.

Since visual speech conveys information about the articulatory gesture, it may strongly relate to speech production skills. Likewise, according to motor and multimodal theories (Fowler, 1986; Liberman & Mattingly, 1985; Massaro, 1987; Schwartz et al., 2012), speech perception relies on the articulatory gestures and visual cues, leading to the prediction that visual speech perception might relate to auditory speech perception. In line with this prediction, fMRI studies have shown that listening to speech and lip-reading visual speech (without any auditory signal) activates the same auditory brain areas in the temporal lobe (Calvert et al., 1997; Campbell et al., 2001; MacSweeney et al., 2000). These activations were found to be specific to speech movements and were not found for meaningless lip movements. In addition to auditory areas, Paulesu et al. (2003) found that lip-reading activated Broca's area, a relevant region for language comprehension. The activation of Broca's area to action observation and language comprehension lead some researchers to propose that it is the human counterpart of the mirror neurons in monkeys (Pulvermüller & Fadiga, 2010). This proposal provided a neural mechanism to combine speech perception and production in the same neural entity in line with the motor accounts of speech perception.

Visual speech information is also used for speech learning. At the same time that infants start to learn the first phonemes, around 4-8 months of age, they focus more on the mouth compared to the eyes (Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015) and are able to match auditory speech with the visual gesture (Pons, Lewkowicz, Soto-Faraco, & Sebastian-Galles, 2009) as well as to use the visual gesture alone to discriminate between different languages (Weikum et al., 2007). In adult populations, the visual gesture has been shown to be helpful to improve L2 phoneme discrimination. Navarra and Soto-Faraco (2007) showed that Spanish dominant Spanish-Catalan bilinguals were able to differentiate between the Catalan /e/-/ɛ/ contrast when presented with audio-visual information but not when presented with just auditory information. Hazan, Sennema, Iba, and Faulkner (2005) showed that training Japanese-English bilinguals to audio-visually discriminate between the /v/-/b/-/p/ labial/labiodental contrast led to greater learning than training the contrast with only auditory information. In this study, audio-visual perceptual training with a less visually distinct gesture, the /l/-/r/ contrast, did not result in any additional benefit for audio-visual training over auditory only training. However, the same audio-visual training led to greater benefits than auditory only training for the production of the /l/-/r/ contrast. These results show benefits of processing visual speech for L2 speech perception and production and suggest that exploiting visual speech cues can improve auditory speech recognition and production ability; thus, individual variability in visual speech recognition may relate to differences in L2 phoneme learning.

In conclusion, speech perception and production models propose a hierarchical structure of sub-lexical and lexical phonological processes. Studying L2 speech perception and production at different levels of the speech processing hierarchy may hinder the observation of the existing relationships between the two skills, explaining the inconsistent correlations reported in previous studies. In addition, the relationship between speech perception and production across the listeners’ languages and sensory modalities remains to be systematically studied.

1.3. The present work

As reviewed, the nature of the relationship between speech perception and production is highly debated. Motor accounts state that speech is perceived through the articulatory gesture and claim for a tight link between speech perception and production (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010). For acoustic accounts of speech perception the acoustic signal is sufficient for perceiving speech (Diehl et al., 2004; Flege, 1995; Hickok & Poeppel, 2007; Klatt, 1979; Kuhl & Iverson, 1995; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994) but speech perception and production might interact in specific contexts. Acoustic models of language learning postulated that production may depend on the phoneme representation created by speech perception (Flege, 1995; Kuhl & Iverson, 1995). The Dual-Stream Model (Hickok & Poeppel, 2007) proposed that speech production may help speech perception only to overcome in difficult listening conditions (Hickok & Poeppel, 2007).

In this dissertation we investigate the nature of the relationship between speech perception and production by exploiting the three different experimental approaches previously reviewed and summarized in Figure 1.2. If speech production is essential for perception, as claimed by motor theories of speech perception, we expect to find mutual influences between speech perception and production regardless of the difficulty of the listening conditions. Furthermore, the relation between speech perception and production should correlate across the listeners’ languages (such as native, L2 and non-native) and sensory modalities (audio and visual), which, to the best of our knowledge, has not been studied previously.

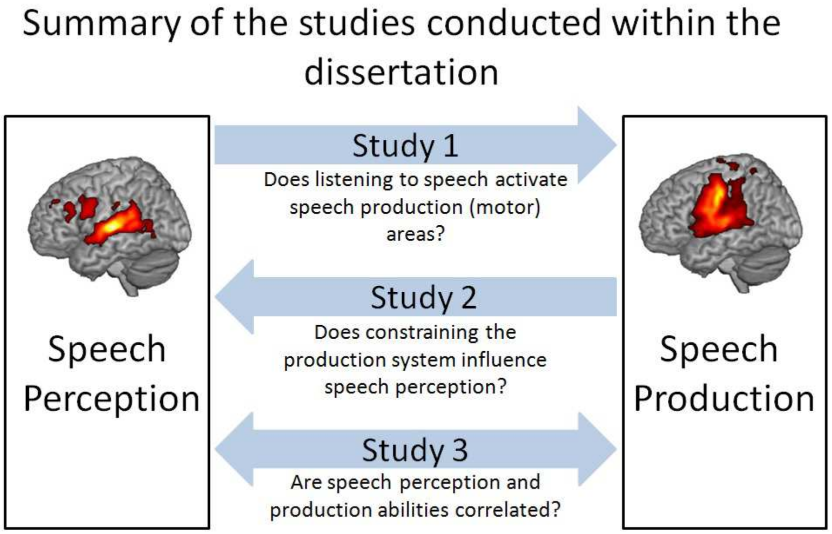

Figure 1.2: Overview of the studies conducted within the dissertation. *Arrows mark the directionality of the relationship between speech perception and production investigated in each of the experiments of the present dissertation: Study 1 analyzes the influence of speech perception on speech production, study 2 tests the influence of speech production on speech perception, and study 3 investigates mutual influences. The research question of each study is detailed below the arrow. The brain areas within the boxes for speech perception and production are from Grabski et al. (2013) and show the brain regions activated in the left hemisphere when native French speakers perceived themselves producing French vowels and when they produced the same French vowels.*

In the first study (Chapter 2) we investigated the influence of speech perception on speech production. Previous studies showed that listening to native speech can activate production areas, such as the primary and pre-motor cortex (Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003; Wilson et al., 2004). However, it remains unclear if these activations are dependent on sensory-motor experience with the phonemes (Swaminathan et al., 2013; Wilson &

Iacoboni, 2006), and if they relate to individual differences in perception and production skills. To test if listening to native and non-native speech activated the production areas differently, we measured corticobulbar excitability of the lip and the tongue muscle when participants listened to native and non-native vowels that varied in their use of the lip and tongue during articulation. Furthermore, we tested if listening to non-native vowels can result in activations specific to the articulatory muscles used during their production. To better understand when the corticobulbar excitability increased, we correlated the excitability with measures obtained from behavioral speech perception and production tasks. During the speech perception task, participants listened to the same vowels presented during the corticobulbar excitability recording and rated the vowels regarding their nativeness and the similarity between pairs of vowels. For the production tasks, participants produced the same vowels presented in the corticobulbar excitability part while the lip activation was measured by means of electromyographic activity (EMG) and tongue activations by means of the first and second formant frequencies, related to tongue height and backness, in the audio signal.

In the second study (Chapter 3) we investigate the influence of speech production on speech perception. Previous studies showed that manipulating the production system influenced speech perception (D’Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Sato et al., 2011) but that the influence could be mediated by task difficulty and/or listening conditions (Alho et al., 2012; D’Ausilio et al., 2012), as well as attention (Möttönen et al., 2013, 2014). We tested if manipulating the production system by means of a somatosensory manipulation, a spoon over the tongue, influenced passive speech perception of vowels differing in tongue height. We tested if attention is necessary for the production system to influence speech perception by recording an electrophysiological measure that is elicited in passive listening conditions – the Mismatch Negativity (MMN). In addition, we investigated if perception difficulties related to the familiarity with the speech sounds modulated the somatosensory influences on speech perception by comparing native and non-native vowels.

In the third study (Chapter 4) we investigated the relationship between speech perception and production across phonological processes (sub-lexical and lexical), listeners’ languages (L1, L2, and non-native), and sensory modalities (auditory and visual). Previous studies are inconclusive in whether a relationship between L2 perception and production abilities exists (Flege et al., 1999; Kartushina & Frauenfelder, 2014; Levy & Law, 2010; Peperkamp & Bouchon, 2011; Rallo Fabra & Romero, 2012). We measured speech perception and production abilities in early and highly proficient Spanish (L1)-Catalan (L2) bilinguals in a variety of speech perception and production tasks. We assessed L2 sub-lexical and lexical speech perception processes by administering the same tasks as Sebastián-Gallés and Baus (2005). We also measured native and non-native phoneme discrimination by means of the MMN (Díaz et al., 2008, in press) and visual speech processing in the native language by means of a cross-modal prediction task developed by Sánchez-García et al. (2011, 2013). The cross-modal prediction task measured benefits of using visual speech to predict the forthcoming auditory signal. We adapted the task to ERP recordings to have complementary information to behavioral measures. To measure speech production, we used an L2 picture naming task (Sebastián-Gallés & Baus, 2005) that required accurate phonological and lexical knowledge. Finally, to investigate the speech-specificity of the relation between speech perception and production, we assessed non-linguistic auditory discrimination and motor skills by means of the MMN elicited by the discrimination of tones and a finger drawing task. We ran factor analysis to investigate how the different measures relate to each other and whether they influence the relation between L2 speech perception and production.

2. STUDY 1

Increased lip corticobulbar excitability during the perception of non-native phonemes

Judith Schmitz [a,b], Eleonora Bartoli [b], Laura Maffongelli [b], Luciano Fadiga [b,c,d], Nuria Sebastian-Galles [a], Alessandro D'Ausilio [b]

[a] Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain
[b] Robotics Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Genova, Italy
[c] Section of Human Physiology, Università di Ferrara, Ferrara, Italy
[d] IIT@UNIFE Center for Translational Neurophysiology, Ferrara, Italy

This article has been submitted for publication to NeuroImage (2015)

2.1. Abstract

Listening to native speech has been shown to activate motor regions, as measured by corticobulbar excitability. In this experiment we explored if the motor regions also recruit during listening to non-native speech, for which we lack both sensory and motor experience. By administering Transcranial Magnetic Stimulation (TMS) over the left motor cortex we recorded corticobulbar excitability of the lip and tongue muscle when Italian participants listened to native-like (/a/,/i/,/u/) and non-native German (/y/) vowels. Results showed highest lip corticobulbar excitability for the non-native and lip articulated vowel /y/. Lip corticobulbar excitability was further correlated with measures obtained in perception and production tasks showing a negative correlation with nativeness ratings and a positive correlation with the uncertainty of lip movement during production of the vowels (standard deviation of lip EMG). These results suggest an active and compensatory role of the motor system during listening to perceptually/articulatory unfamiliar phonemes.

2.2. Introduction

Listening to speech activates temporo-parietal regions, as well as the motor system. Activations of the motor regions, including the representation of articulatory muscles of the primary motor cortex, has been tested via corticobulbar excitability modulations (D’Ausilio, Jarmolowska, Busan, Bufalari, & Craighero, 2011; Fadiga et al., 2002; Murakami et al., 2011; Rogers, Möttönen, Boyles, & Watkins, 2014; Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Sato, Buccino, Gentilucci, & Cattaneo, 2010; Sundara et al., 2001; Watkins et al., 2003). Furthermore, the application of (repetitive) Transcranial Magnetic Stimulation (rTMS) to the premotor (Grabski, Tremblay, Gracco, Girin, & Sato, 2013; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Sato, Tremblay, & Gracco, 2009) or the primary motor cortex (Bartoli et al., 2015; D’Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011; D’Ausilio et al., 2009, 2012; Möttönen et al., 2013, 2014; Möttönen

& Watkins, 2009; Rogers et al., 2014) showed that the motor system may exert a causal modulatory role in both the discrimination and identification of speech sounds. These results are usually interpreted in line with theories for which knowledge of articulatory gestures may be re-used during speech identification and discrimination tasks (Liberman et al., 1967; Pulvermüller & Fadiga, 2010).

While the studies described so far show motor processes during the analysis of native speech, less is known about non-native phonemes, which are biomechanically possible oral configurations for which we lack both sensory and motor experience. Neuroimaging studies show that, besides temporo-parietal regions, part of the motor system (premotor cortex) is activated during both listening and production of native phonemes (Wilson et al., 2004) and listening to non-native phonemes (Wilson & Iacoboni, 2006). These latter results extend this evidence to African clicks, that are judged as non-speech sounds in other languages (Agnew, McGettigan, & Scott, 2011). Regarding the primary motor cortex, a TMS study by Swaminathan et al. (2013) measured lips corticobulbar excitability while native and second language learners of English were viewing known speech (English), unknown speech (Hebrew), non-speech movements (gurns) or a static face. Results showed that, in both native and second language learners, the motor cortex is engaged more strongly when viewing faces producing known speech.

However, visual speech differs from auditory speech in that it has less discriminative value. For example, /d/, /t/, /g/ and /k/ sounds map into very similar visual gestures or visemes. In fact, auditory identification of the /d/ sound will not be affected by the parallel presentation of /t/, /g/, /k/ visemes (MacDonald & McGurk, 1978). Furthermore, the reduced informational value of visemes also interacts with language proficiency. In this regard, discriminating if two sentences are spoken in the same language or in two different languages can easily be performed in the auditory modality. In the visual modality such task can only be performed if at least one of the languages is either native (Soto-Faraco et al., 2007) or a high proficiency has been achieved with it (Swaminathan et al., 2013; Weikum et al., 2013). Therefore, measuring motor activities in native versus non-native speech, by using visual speech material, may be a suboptimal choice. In fact, language experience as well as the specific choice of visemes may render the discrimination impossible for the non-native speaker. In this sense, it would be difficult to understand if corticobulbar modulations are driven by a general preference for native speech or it is due to a better discriminability of native speech when presented in the visual modality (Soto-Faraco et al., 2007; Swaminathan et al., 2013; Weikum et al., 2013).

In this experiment we investigated if the motor system activities during auditory speech listening depend on sensorimotor experience with the phonemes. We tested this hypothesis by measuring lip and tongue corticobulbar excitability during passive listening to native and non-native vowels. Stimuli consisted in German vowels (/a/, /i/ and /u/) having a native counterpart in the language of the participants (Italian). The non-native vowel (/y/) has some articulatory similarities to native vowels (similar to /i/ for the front-high tongue position and to /u/ for lip rounding), but it is not familiar with our subjects' repertoire. Furthermore, we tested for correlation between corticobulbar excitability and subjective ratings regarding vowels nativeness (for each vowel) and similarity (between each pair of vowels), as well as participant's vowel production characteristics. These measures included lip muscle electromyographic activity (EMG) and the acoustic signal related to tongue height (first formant frequency, F1) and tongue backness (second formant frequency, F2).

The exploration of motor activities by measuring corticobulbar excitability let us formulate different hypotheses in this regard. In fact, corticobulbar excitability could scale for the amount of sensorimotor experience (hypothesis 1). In this case, listening to over-trained (native) speech sound should elicit stronger motor responses, in agreement with the previous report on visual speech perception (Swaminathan et al., 2013). Otherwise, corticobulbar excitability could instead be larger while listening to untrained speech sounds. The lack of an acoustic-motor model for non-native speech sounds (hypothesis 2) might instead promote additional motor compensatory activities.

2.3. Methods

2.3.1. Participants

17 native Italian speakers (9 female, mean age 23.59±4.81 years) took part in this study after giving informed consent, according to the Declaration of Helsinki and to the recommendations of the local Ethical Committee ASL-3 ("Azienda Sanitaria Locale"- Local Health Unit, Genoa, Italy) authorizing the protocol. All participants had normal hearing, were right-handed (Oldfield, 1971) and did not report any neurological/psychiatric disease. They were Italian university students, had no professional training in phonetics and were not proficient in any language whose phonological repertoire includes the non-native vowels used in the experiment, such as French and German (as accessed by a language questionnaire). One participant was excluded in the lip recording and one in the tongue recording due to technical issues. One further participant was excluded in the similarity ratings task due to the misunderstanding of the task. Two further participants were removed from all tasks in relation to outlier behavior in one of the tasks (i.e., the participants rated the foreign German vowel /y/ higher in nativeness with respect to the German vowel /u/, that is present in the Italian phonological repertoire). The final sample included 14 participants for the lip and tongue recordings, 14 for the correlations with the nativeness ratings, 13 for the correlations with the similarity ratings, and 14 for the correlations with the speech production data.

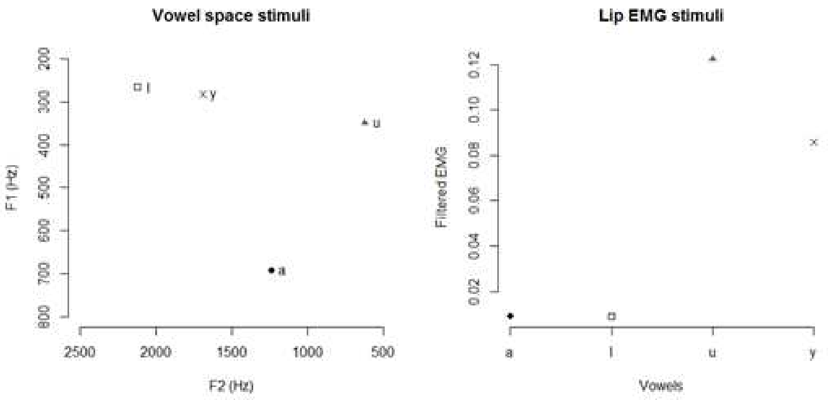

Figure 2.1: *Stimuli. Vowel space (left) and lip EMG (right) of the stimuli.*

2.3.2. Stimuli

The seven German vowels: /a/, /e/, /i/, /o/, /u/ (having a counterpart in Italian) and /ö/ and /y/ (unknown to Italians) were recorded by a male native German speaker using a microphone and surface electrodes placed on the lower and upper right side of the lip to record the electromyography (EMG) of the *orbicularis oris* (OO) muscle. Each vowel was recorded twelve times and the best exemplar for each vowel, matching in pitch (127 Hz) and intensity (75 dB), was selected. To equal the length for the stimuli, three-hundred ms were selected from the steady middle part of the vowel and 25 ms cosines onset and offset were applied using Praat software (Boersma & Weenink, 2010). Values for the first and second formant frequencies and the mean lip EMG (band-pass filtered between 20-250 Hz and low pass filtered with a cut-off frequency of 3 Hz) for the selected stimuli /a/, /i/ /u/ and /y/ are shown in Figure 2.1 (F3 values: /a/: 2371.28, /i/: 2905.41, /u/: 2081.75 and /y/: 1999.02). The mid-vowels /e/, /o/ and /ö/ were removed from the final analysis because the vowels /e/ and /o/ that were expected to have a counterpart in Italian, did not show the expected nativeness ratings (see procedure speech perception task for the results).

2.3.3. General procedure and statistical analysis

Participants were first asked to fill-in a language questionnaire to test their degree of exposure to the non-native vowels (by asking for knowledge in foreign languages and dialects). The experimental session consisted of three parts: a TMS experiment, a speech perception task and a speech production task. The TMS part was always run as first, while the order of the behavioral perception and production parts was counterbalanced across participants. All experiments were programmed using Psychtoolbox functions (Brainard, 1997; Pelli, 1997), running on MATLAB ® (The MathWorks Inc., Natick, MA). For the analysis, all data were pre-processed using MATLAB ® (The MathWorks Inc., Natick, MA). Statistical analyses (ANOVAs, t-tests and correlations) were performed by means of R statistical software (R Core Team, 2013). All t-tests were corrected using the Bonferroni correction.

2.3.3.1. Speech perception tasks

The perception tasks consisted of a nativeness and a similarity rating task.

In the nativeness rating task, each vowel (/a/, /e/, /i/, /o/, /u/, /ö/, /y/) was rated three times (total: 21 trials) in a random order on a visual analog scale ranging from “poco” (meaning “less”) on the left side of the screen to “tanto” (meaning “more”) on the right side of the screen. Answers were given by clicking with the mouse on any location along the continuum. For the analysis, the nativeness ratings were averaged for each vowel for each participant. An ANOVA showed a main effect for the within-subject factor “vowel” ($F(6,84)=36.610$, $p<0.001$). Follow-up t-tests showed that, as expected, the vowels /ö/ (21.667±3.670%) and /y/ (14.000±1.966%) were rated significantly less native compared to all other vowels (/a/: 92.578±1.777%, /e/: 62.089±4.427%, /i/: 87.156±2.336%, /o/: 59.200±4.712%, /u/: 75.044±4.330%, all $p<0.01$). Unexpectedly, the mid-vowels /e/ and /o/ were also rated less native compared to the most native rated vowel /a/ (/e/-/a/: $t(14)=4.243$, $p=0.017$, /o/-/a/: $t(14)=4.568$, $p=0.009$), suggesting that the mid-vowels were poor exemplars of the Italian mid-vowels. To keep the most extreme native and non-native vowels, all mid vowels were excluded (/e/, /o/ and /ö/), leaving for the following analyses the vowels: /a/, /i/, /u/ and /y/.

The similarity rating task was identical to the nativeness-rating task, only that pairs of vowels were played (with an interval of 3.5 seconds) and rated. From the total of 42 trials (twice each pair with the order of the vowels counterbalanced), similarity of the selected non-native vowel /y/ to the selected native vowels (i.e. /y-a/, /y-i/, /y-u/) was computed and averaged for each participant. An ANOVA with the within-subject factor “vowel similarity” showed a main effect ($F(2,26)=10.870$, $p<0.001$). Follow-up t-tests showed only a significant difference between the similarity /y-a/ (21.25±5.682%) and /y-u/ (61.500±5.278% t(13)=-4.117, p=0.004) but not for /y-i/ (43.143±5.513%) with any other similarity (all p>0.05).

2.3.3.2. Speech production task

Audio and EMG activity from the OO muscle were recorded from each participant during the production task. Participants produced each vowel after seeing the German grapheme (<a>, <e>, <i>, <o>, <u>, <ö>, <ü>) corresponding to the vowel (/a/, /e/, /i/, /o/, /u/, /ö/, /y/) on the screen. We recorded 8 repetitions of each vowel and the last five were subsequently used for the analysis. Previous to the production trials, participants were familiarized with the writing of the German graphemes. They were presented three times with each German grapheme followed by the playback of each vowel (<ö> and <ü> are not used in standard Italian).

For the analysis, 300 ms were selected from the central part of the recorded audio and EMG. In the audio data, first and second format frequencies (F1 and F2) were obtained using Praat (Boersma & Weenink, 2010). The EMG data was band-pass filtered between 20-250 Hz and then low pass filtered with a cut-off frequency of 3 Hz. EMG for each trial was rectified and integrated, obtaining in this way a measure of the area under the curve, which was standardized for each participant using z-scores. Average (mean) and standard deviations (SD) were calculated for each vowel and participant. ANOVAs including the within-subjects factor selected vowel (/a/, /i/, /u/, /y/) were performed for each measure for the means and SD.

This was firstly done in order to confirm that the production was as expected for native vowels and secondly to gain information about the accuracy and variability due to the production of the non-native vowel.

Regarding the mean values, F1 and F2 productions (Figure 2.2) show the expected vowel space, similar to the vowels recorded by the native speaker for the stimuli (Figure 2.1). F1 (tongue height) mean values were all different, expect for /i/ and /y/ (F(3,42)=271.400, p<0.001, t-tests all p<0.01, except /i/-/y/). F2 (tongue backness) mean values were all different, except /a/ and /y/ (F(3,42)=90.300, p<0.001, t-tests all p<0.01, except /a/-/y/), indicating that /y/ was "central", lying in between /i/ and /u/, while for the native speaker /y/ was closer to /i/ than /u/ (see Figure 2.1). Mean lip EMG (Figure 2.3) showed the expected difference between rounded (/u/, /y/) and unrounded (/a/, /i) vowels (F(3,39)=43.670, p<0.001, t-tests all p<0.01, except /a/-/i/ and /u/-/y/), similar to lip EMG for the native speaker for the stimuli (Figure 2.1).

The SD for F1 and F2 (Figure 2.2) showed no difference for F1 (F(3,42)=2.051, p=0.121) but for F2 (F(3,42)=15.63, p<0.001), with significant differences for /y/ with all other vowels (all p<0.05), indicating higher variability for the non-native vowel for F2. The ANOVA for the SD of the lip EMG (Figure 2.3) showed significant differences for /a/ with the lip articulated native /u/ and non-native /y/ (F(3,39)=5.798, p=0.002, t-tests /a/-/u/ and /a/-/y/ p<0.01), thereby showing the highest variability for both the native and non-native rounded vowels (/u/, /y/).

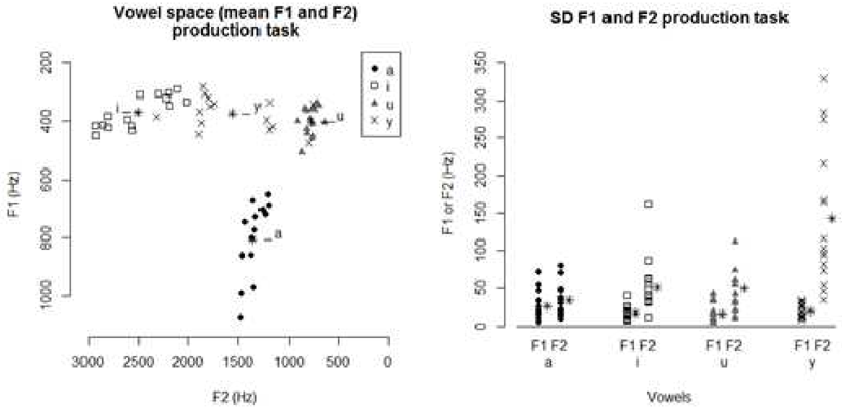

Figure 2.2: *Production task. Left: vowel space representing the mean value for F1 and F2 for each participant and vowel. Right: SD F1 and F2 for each participant and vowel. Stars mark the mean values for all participants.*

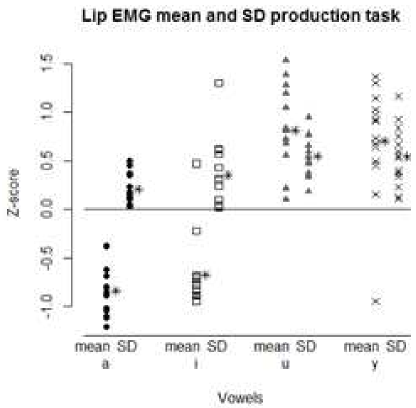

Figure 2.3: *Production task. Lip EMG mean and SD for each participant and vowel. Stars mark the mean values over all participants.*

2.3.3.3. TMS experiment

Two surface EMG electrodes were placed on participants’ upper and lower side of the right lip (*orbicularis oris*, OO) and two additional EMG electrodes were glued (Hystoacryl, B. Braun Surgical SA) on the tip and middle part of the right side of the tongue muscle (dorsal surface). MEPs were recorded by means of a wireless EMG system (Aurion, ZeroWire EMG; 2KHz sampling). TMS was delivered through a figure-eight coil (70 mm) and a Magstim 200 monophasic stimulator (Magstim Co., Whitland, UK).

The TMS experiment was divided into training phase, cortical mapping and experimental part. During the training phase, participants learned how to maintain a 1.5-2 seconds sustained contraction of both lips and tongue, corresponding to 30% of the maximal EMG activity. A yellow dot appeared on the screen to signal the beginning of the contraction, followed (random interval of 1-1.5 seconds) by a beep. The disappearance of the yellow dot (200 ms after the beep) signaled the end of the trial. Lips were contracted by rounding and protruding them. The tongue was contracted by pressing it against the upper front palate. Participants could see the EMG activity of the two muscles on the screen and received feedback from the experimenter if necessary. When a satisfactory contraction was achieved, the cortical mapping part started. The hot spot was identified during contraction of both muscles. Scalp position, coil orientation and intensity of stimulation were optimized to obtain the lowest possible intensity to elicit a reliable and repeatable MEP on both muscles. The criterion was to produce a MEP of at least 200 microvolts in both muscles on 5 consecutive trials. Once the criterion was met, location was marked on the scalp and the coil position was fixed by a mechanical support and was continuously monitored by the experimenter.

At the beginning of each experimental trial a yellow dot appeared on the screen to signal the contraction of the lip and tongue muscle. After a random interval of 1-1.5 seconds one of the seven vowels was played. The TMS pulse was triggered at 150 ms after stimulus onset. The delay between trials ranged between 5-5.5 seconds, giving a total time of 7 seconds for each trial. The presentation of auditory stimuli was pseudo-randomized so that the presentation of the same vowel was equally spread throughout the experiment and could never appear twice in a row. Each vowel type was repeated 14 times, 10 times with TMS and 4 times without TMS. TMS and no-TMS trials were randomized. The whole experiment consisted in 98 trials and it was divided in 2 blocks of 49 trials each. An additional task was added in order to keep participants engaged throughout the experiment. After the end of 12 random trials we asked participants if a certain native vowel, printed on the screen, was the same as the one they just listened to (one-back task). Responses were given by pressing one of two buttons on a keyboard, with their right hand. The task was equally spread during the experiment and it was never presented twice in a row.

In the analysis of MEP size, we computed the area under the rectified curve for each trial and for each muscle separately. For each participant, single trials were removed if the MEP area exceeded 2 standard deviations from the average MEP area or if the muscle contraction was above 2 standard deviations from average muscle contraction in a time window just prior to the TMS pulse delivery (-100 to 0 ms). The MEPs were then standardized for each participant using z-scores and the trials corresponding to the same vowel were averaged together. The ANOVA was run separately for the lip and tongue MEPs testing for differences in MEP size accounted by the within-subject factor “vowel” (selected vowels: /a/, /i/, /u/, /y/). Significant effects (main effect of “vowel” with $p<0.05$) were further analyzed with t-tests.

In case of significant main effects, Pearson correlations between the MEPs (for each vowel and participant) and the other measures extracted from perception and production tasks were carried out. Only for the similarity ratings, MEPs were standardized with respect to the non-native vowel /y/ to have comparable measures for both tasks. For the production data, we planned to compare lip and tongue MEPs with the production data corresponding to the same muscle. Lip MEPs were correlated with the mean and SD of lips EMG and audio F2 (related to lip rounding). Tongue MEPs were correlated with the mean and SD of audio F1 (related to tongue height) and F2 (also related to tongue backness).

2.4. Results

2.4.1. Corticobulbar excitability

The ANOVA for the tongue MEPs showed no significant main effect of “vowel” (F(3,39)=1.557, p=0.215). The ANOVA for the lip MEPs showed a significant main effect of “vowel” (F(3,39)=4.074, p=0.0131). Follow-up Bonferroni corrected t-tests revealed a difference in MEP size between the vowels /a/ and /y/ (t(13)=-3.348, p=0.031, see Figure 2.4). To further explore the significant differences for the lip MEPs related to the listening of the different vowels, correlations measures were performed between the lip MEPs and the scores obtained in the different perception and production tasks.

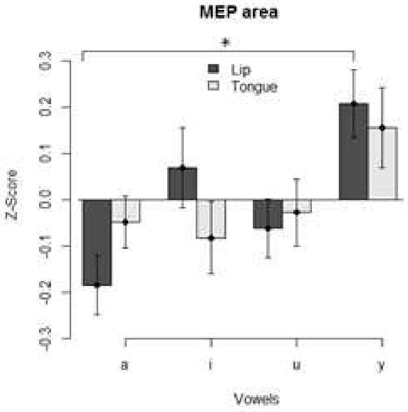

Figure 2.4: *TMS results. MEP area +/- SEM for the lip (dark) and tongue (light) muscle for the German vowels /a/, /i/, /u/ (having a counterpart in Italian) and /y/ (non-native).*

2.4.2. Correlation between corticobulbar excitability and the behavioral tasks

The correlation between the lip MEPs with the perception data revealed a significant negative correlation with the nativeness ratings (r=-0.373, t(54)=-2.950, p=0.005, see Figure 2.5). Thus, higher MEPs size was associated to lower nativeness ratings. On the contrary, no correlation could be found between lip MEPs and the similarity ratings (r=0.1001, t(37)=0.6156, p=0.542).

A positive correlation was present between the lip MEPs and the SD of lip EMG only (r=0.420, t(54)=3.40, p=0.001, see Figure 2.5), whereas other production-related measures (mean EMG, SD F2 and mean F2) did not yield to significant correlation results (all p>0.05).

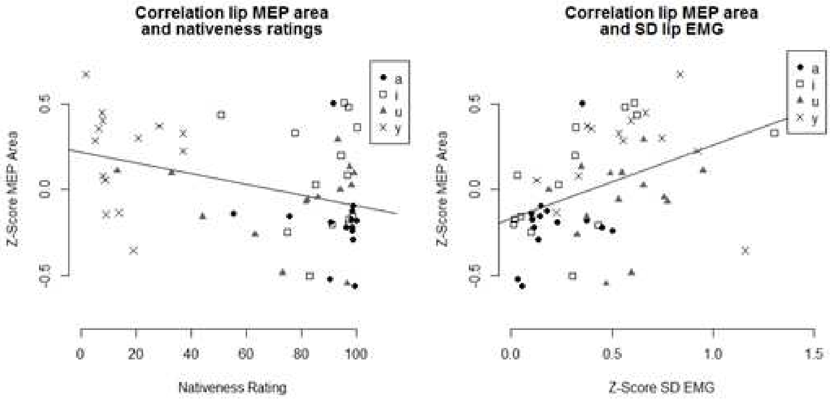

Figure 2.5: *Correlations for the lip MEP area with left: nativeness rating (r=-0.373) and right: SD of the lip EMG (r=0.420).*

2.5. Discussion

In this study we investigated if activations of the motor system during speech listening depend on sensorimotor experience with the phonemes. In line with hypothesis 2, which stated that the lack of acoustic-motor model for the non-native speech sounds might lead to motor compensatory activities, the strongest lip facilitation was found for the non-native vowel /y/ compared to the native vowel /a/. The other native vowels /i/ and /u/ were in between the native vowel /a/ and the non-native vowel /y/, probably because, in contrast to /a/, the native vowels /i/ and /u/ are produced by lip movements (/i/ lips pulled back and /u/ lip rounding), leading to some modulation of corticobulbar excitability that is smaller than for /y/, which is both non-native and is articulated with a lip movement.

Further support for hypothesis 2 came from correlations of the lip MEPs with both the speech perception and production data. A negative correlation between the nativeness ratings and the lip MEPs were found, showing that lip corticobulbar excitability increases when the perceived nativeness decreases. Additionally, a positive correlation between the lip MEPs and the standard deviation of the lip EMG was found, showing that lip corticobulbar excitability increases when the production of the vowels was more variable and thus participants were more insecure about how much movement of the lip was expected to produce the vowel. Taken together, these results suggest that corticobulbar excitability of the lip is highest for perceptually and articulatory unfamiliar vowels, in line with the idea that the motor cortex might compensate for not having an acoustic-motor model for the non-native speech sounds.

This interpretation is in line with a previous neuroimaging study reporting higher activations in the motor system during passive listening to non-native phonemes compared to native ones (Wilson & Iacoboni, 2006). Additionally, the motor system has been shown to increase its activation during the identification of a difficult second language contrast (/l/-/r/ for Japanese) in comparison to both an easy second language contrast (/b/-/g/ for Japanese, Callan et al., 2003) and the same contrast (/l/-/r/) in native speakers (English) for which the contrast is easier to identify (Callan, Jones, Callan, & Akahane-Yamada, 2004). Based on these results, Callan et al. (2003, 2004) proposed that second language learners do not only reorganize auditory brain regions when learning a new difficult second language contrast, but that they additionally engage more strongly brain networks connecting articulatory-auditory and articulatory-orosensory brain regions. In parallel to studies showing motor compensatory activations during the perception of non-native phonemes, lip MEPs were also found to increase when raising the difficulty level in native speech perception (Murakami et al., 2011). This suggests that motor compensatory activities can be found when the audio signal alone could not be sufficient to identify a perceived phoneme in a variety of different tasks, including difficult native, second language and non-native contrasts.

The present results are however at odds with a recent study by Swaminathan et al. (2013) showing the opposite pattern of corticobulbar excitability for visual speech. In fact the authors found higher corticobulbar activity of the lip muscle for viewing a known language compared to viewing an unfamiliar language. This is likely due to the very different stimuli. Visual continuous speech as used in Swaminathan et al. (2013) is very difficult to understand without training, even in the native language (Bernstein, Demorest, & Tucker, 2000; Ronquest, Levi, & Pisoni, 2010; Soto-Faraco et al., 2007). Gaining information about articulatory information from visual continuous speech in a non-native language might thus just have been impossible. In contrast, auditory-presented single vowels are easily identified in the native language, and might therefore also be more informative for non-native vowels. Indeed, our study showed that participants were able to gain articulatory related information from the speech signal not only by increased lip corticobulbar excitability during the perception task, but also by imitating the non-native vowel relatively well in the production task, placing it close to the values of native speakers.

In this study we simultaneously measured corticobulbar excitability of the lip and tongue muscle but we only found differences for the lip muscle. Previous studies (Fadiga et al., 2002; Watkins et al., 2003) using native speech listening, found MEP enhancements for both muscles (though in separate experiments) when viewing speech, listening to speech and listening to phonemes articulated with the muscle that was recorded. The lack of tongue modulation in our study might be related to the lip movement being more salient for the non-native vowel /y/. Previous research has shown that although front rounded vowels, such as /ö/ and /y/ are more similar in the acoustic space and tongue position to front vowels such as /e/ and /i/ respectively, American English listeners (a language that, as Italian, does not possess front rounded vowels) assimilated them to back rounded vowels, /o/ and /u/ respectively (Strange, Levy, & Law, 2009). In fact, they had more difficulty in discriminating front rounded vowels from back ones (e.g /y-u/ compared to /y-i/) even when a high proficiency in the foreign language was acquired (Levy, 2009). Interestingly, larger lip MEPs to the non-native vowel /y/ shows that participants were able to extract information about the gesture of the non-native phoneme – a lip movement – probably by assimilating the non-native vowel to the articulatory similar native vowel /u/ that also uses lip rounding and thus actively trying to find a match for the perceived phoneme. In support of this possibility, the subjective similarity ratings between the vowel /y/ and the other native vowels showed a decrease in ratings from /u/ (61.500 ± 5.513%) to /i/ (43.143 ± 5.278%) to /a/ (21.250 ± 5.682%).

Although newborns are able to perceive most phoneme contrasts from any language at birth (Werker & Tees, 1984), this capacity rapidly declines at the end of first year of life, as infants start to focus on the phonemes of their native language (Werker & Tees, 1984). The ability to successfully acquire later in life the sounds of a new language largely depends on the relative structure of the phonological system of our native language and the foreign language (Best & Tyler, 2007; Best, 1995; Flege, 1995, 2003). Phonemes that fall within the same phonological category in the native language but in different ones in the foreign language – such as the /l/-/r/ contrast for Japanese learners of English or /e/-/ɛ/ for Spanish learners of Catalan – are extremely hard to acquire (Miyawaki et al., 1975; Pallier et al., 1997), due to the assimilation of the new sounds to the native ones. Our study shows that even when confronted to an unknown and untrained phoneme, the corticobulbar excitability of the lip muscle – a muscle used during the articulation of the non-native phoneme – increases. This is probably due to an assimilation mechanism which compensates for the lack of an acoustic-motor model for the non-native speech sounds. These results suggest that the motor system plays an active role in speech perception, even when confronted with new and untrained phonemes and this role might be relevant even during the first contact with a new, foreign language.

Acknowledgements

This research was supported by the European Community Grant POETICON ++ (STREP-Project ICT – 288382) awarded to L.F, and grants from the European Commission Seventh Framework Programme (FP7/2007-2013): ERG grant agreement number 323961 (UNDER CONTROL); Cooperation grant agreement number 613465 – (AThEME), the Spanish Ministerio de Economía y Competitividad (PSI2012-34071) and the Catalan Government (SGR 2014–1210) awarded to N.S.G.. She also received the prize

‘‘ICREA Acadèmia’’ for excellence in research, funded by the Generalitat de Catalunya.

2.6. References

Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience*, *23,* 4038–47.

Bartoli, E., D’Ausilio, A., Berry, J., Badino, L., Bever, T., & Fadiga, L. (2015). Listener-Speaker Perceived Distance Predicts the Degree of Motor Contribution to Speech Perception. *Cerebral Cortex*, *25,* 281–288.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62,* 233–252.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204). Baltimore: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing Company.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program].

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10,* 433–436.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, *22,* 1182–94.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*, 113–124.

D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, *118*, 9–14.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, *48*, 882–887.

D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: priming and coarticulation effects. *Neuropsychologia*, *49*, 3670–3676.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*, 399–402.

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Baltimore: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–355). Berlin: Mouton de Gruyter.

Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain Research*, *1515*, 55–65.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The*

Journal of the Acoustical Society of America, *126*, 2670–2682.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*, 253–257.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*, 1692–6.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of Linguistic Experience: The discimination of [r] and [l] by native Speakers of Japanes and English. *Perception & Psychophysics*, *18*, 331–340.

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, *23*, 1190–1197.

Möttönen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *The Journal of Neuroscience*, *34*, 4064–4069.

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience*, *29*, 9819–9825.

Murakami, T., Restle, J., & Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia*, *49*, 2045–54.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*, B9–B17.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pulvermüller, F., & Fadiga, L. (2010). Active perception:

sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*, 351–360.

R. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rogers, J. C., Möttönen, R., Boyles, R., & Watkins, K. E. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology*, *5*, 754.

Ronquest, R. E., Levi, S. V, & Pisoni, D. B. (2010). Language identification from visual-only speech signals. *Attention, Perception & Psychophysics*, *72*, 1601–1613.

Roy, A. C., Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *Journal of Physiology*, *102*, 101–5.

Sato, M., Buccino, G., Gentilucci, M., & Cattaneo, L. (2010). On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Communication*, *52*, 533–541.

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*, 1–7.

Soto-Faraco, S., Navarra, J., Weikum, W., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, *69*, 218–231.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naive American listeners. *The Journal of the Acoustical Society of America*, *126*, 1461–1476.

Sundara, M., Namasivayam, A. K., & Chen, R. (2001). Observation-execution matching system for speech: a magnetic stimulation study. *Neuroreport*, *12*, 1341–1344.

Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., & Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain and Language*, *126*, 1–7.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994.

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience*, *7*, 1–8.

Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life, *7*, 49–63.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage*, *33*, 316–325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702.

3. STUDY 2

Attention modulates somatosensory influences in passive speech listening

Judith Schmitz [a], Begoña Díaz [a], and Nuria Sebastian-Galles [a]

[a] Center for Brain and Cognition, Department of Technology, University Pompeu Fabra, Barcelona, Spain

This article has been submitted for publication to Frontiers in Psychology - Language Science (2015)

Schmitz, J, Díaz B, Sebastian-Galles N. Attention modulates somatosensory influences in passive speech kistening. J Cogn Psychol (Hove, England). 2016. 28; 791-806. DOI: 10.1080/20445911.2016.1206.107

4. STUDY 3

Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities

Judith Schmitz [a], Begoña Díaz [a], Karla Fernández Rubio [a], and Nuria Sebastian-Galles [a]

[a] Center for Brain and Cognition, Department of Technology, University Pompeu Fabra, Barcelona, Spain

This article has been submitted for publication to Cognition (2015).

Schmitz, J, Díaz B, Fernández Rubio K, Sebastian-Galles N. Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. Language, Cognition and Neuroscience. 2017; 21 Oct: 1-20. DOI: 10.1080/23273798.2017.1390142

5. DISCUSSION

The main goal of this dissertation was to investigate the nature of the relationship between native and non-native speech perception and production. To this aim, we conducted three experiments following different research approaches and methodologies. In the first study we investigated the influence of speech perception on production. Based on previous studies showing that listening to native speech activates production areas, in the first study we tested if listening to non-native speech results in similar activations of the production system. To this aim we recorded corticobulbar excitability while participants listened to native and non-native phonemes and related the excitability to individual data obtained through speech perception and production tasks. In the second study we investigated the influence of speech production on perception. We examined if constraining the production system by means of a somatosensory manipulation, a spoon over the tongue, influenced speech perception in passive listening and whether this influence was dependent on the listeners' familiarity with the phonemes. In the third study we analyzed the correlation between speech perception and production abilities. We measured speech perception and production abilities in early Spanish (L1)-Catalan (L2) bilinguals. Speech perception was assessed for a variety of tasks covering different phonological processes (sub-lexical and lexical), listeners' languages (L1, L2 and non-native), and sensory modalities (auditory and visual).

The three research studies presented in this dissertation aimed at contributing to the ongoing debate about the nature of the relationship between speech perception and production: Whether speech perception and production share the intended articulatory gestures as the basic processing unit and are closely related, as proposed by motor views, or rather if speech perception is an auditory process that does not need the production system, at least in normal listening circumstances. In the following, we first summarized the results of each study and then related the overall results to the current literature. To conclude, we outlined open questions and suggested future lines of research.

5.1. Summary and discussion of the results

5.1.1. Increased lip corticobulbar excitability during the perception of non-native phonemes

In the first study (Chapter 2) we tested if the motor activations found while listening to native speech are dependent on former sensory-motor experience with the phonemes. To this aim we recorded corticobulbar excitability of the lip and tongue muscle while Italian participants listened to German vowels that either had a counterpart in the native language of the participants, in particular the vowels /a/, /i/ and /u/, or were non-native to the participants, the vowel /y/. Additionally, we correlated corticobulbar excitability with measures obtained in behavioral speech perception and production tasks. For speech perception, we recorded nativeness ratings of the vowels and similarity ratings between pairs of vowels. For speech production, we recorded the lip EMG and the acoustic signal that gives a measure of tongue height (F1) and tongue backness (F2). Two speech production measures were calculated: the mean value, related to the accuracy of the production, and the standard deviation, related to the uncertainty of the use of the muscle during articulation.

Results showed that lip corticobulbar excitability increased for a combination of lip use during articulation and non-nativeness. Highest excitability was found for the non-native and lip rounded vowel /y/, intermediate excitability was found for the native vowels /i/ and /u/ that use a lip movement during articulation (lips pulled backwards or rounded respectively), and the lowest excitability was found for the native vowel /a/ that does not recruit the lips during articulation, with significant differences between the extremes (/y/-/a/). Correlations further showed that lip corticobulbar excitability was negatively correlated with the nativeness ratings and positively with the uncertainty of the lip movement during articulation (measured by the standard deviation of the lip EMG). No differences in corticobulbar excitability were found for the tongue. We suggested that this is due to the more salient lip-rounding for the vowel /y/, which has been shown in previous studies of English learners to be assimilated more often to native lip-rounded vowels than to native vowels with a similar tongue height (Strange et al., 2009). The results showed that lip corticobulbar excitability increases for perceptually and articulatory unfamiliar vowels. We interpreted the enhanced excitability of the motor system as a compensatory mechanism that supports speech perception in absence of previous sensory-motor experience with the vowels.

Our results are in line with previous neuroimaging studies reporting higher motor activations when listening to non-native speech (Wilson & Iacoboni, 2006), a difficult L2 contrast (Callan et al., 2003), and degraded native speech (Murakami et al., 2011). Together, these studies suggest that the motor cortex is actively participating in speech perception by trying to identify the articulatory gestures of the perceived phonemes, even when they are unfamiliar. We interpret the results as evidence that speech perception recruits motor cortices to compensate for the lack of an accurate auditory-motor model of the non-native phonemes. Our study adds to the literature by showing that these compensatory mechanisms can already be found at initial encounters with a new language and are related to the non-nativeness of the vowels, both perceptually and articulatory. The results support the proposal that speech production is involved in perception. However, as argued in the general introduction, influences of speech production on perception, such as the here presented, are not sufficient to determine whether speech production is essential to perception, as claimed by the motor view of speech perception (Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010), or it is a compensatory mechanism to overcome difficulties in perception, as postulated by the Dual-Stream model (Hickok & Poeppel, 2007) within the acoustic view of speech perception.

5.1.2. Attention modulates somatosensory influences in passive speech listening

In the second study (Chapter 3) we aimed to investigate the specificity of production influences in speech perception. We tested if constraining the production system by means of a somatosensory manipulation on an articulator, a spoon over the tongue, modulated speech perception in passive listening conditions, by measuring the electrophysiological response MMN. We tested if this influence is articulatory-dependent by measuring the MMN for vowels varying in tongue height and whether the influences depended on the familiarity with the vowels, by recording the MMN for native vowels and non-native vowels. The standard stimulus, repeated frequently, was the native vowel /e/ articulated with a front mid tongue height. The deviants, presented infrequently, were grouped based on the combination of nativeness and tongue height, resulting in four deviant groups: native-high (/i/ and /u/, front and back respectively), native-low and mid (/a/ and /o/, central and back), non-native-high (the Finnish vowel /y/, mid-high), and non-native-mid (the Finnish vowel /ö/, central). The same participants heard all the vowels with and without the spoon over the tongue (a spoon and a no-spoon block) with the order of the blocks counterbalanced across participants.

Unexpectedly, results showed no effect of the spoon, nativeness, or any interaction between the variables. Instead, the order of the spoon conditions influenced the MMN amplitude: Starting the experiment with the spoon over the tongue resulted in higher MMN values as compared to starting with the no-spoon block. We proposed that the block order effect might be caused by an increase of participants' attention or alertness because of the motor constrain (having a spoon in the mouth) that lasted throughout the experiment, even when participants no longer held the spoon over the tongue. This hypothesis was tested in a second experiment in which participants started the experiment with a somatosensory manipulation but applied to a non-articulator, a spoon in the hand, while listening to the same speech stimuli than in the first experiment. Results showed an increased MMN amplitude for participants starting with the spoon in the hand that was not different from starting with the spoon over the tongue. Both MMN amplitudes for starting with the spoon were higher compared to starting with the no-spoon block. The results suggested that the MMN was influenced by the alerteness or attention of the participants when an irrelevant task for speech perception, holding a spoon over the hand or the tongue, is performed at the beginning of the study.

The current results did not show any influences of the somatosensory manipulation, the spoon over the tongue, in passive listening condition. Yet, they add important methodological information about the factors that influence the MMN and how somatosensory manipulations impact the experimental situation. Our results showed that the MMN elicited by speech sounds is influenced by attention when participants start the experiment performing an additional task, holding a spoon over the tongue or in the hand. This study thereby shows that somatosensory manipulations applied during passive listening tasks engage attention and caution for the use of somatosenory manipulations in passive listening conditions.

With the current study the question of whether speech production contributes to perception in passive listening conditions remains unclear. On the one hand, our results can be interpreted to show no influences of somatosensory manipulations in passive speech perception. On the other hand, it is possible that the attentional effect induced by starting the experiment with the somatosensory manipulation diluted any effect of the somatosensory manipulation on speech perception.

5.1.3. Exploring the relationship between speech perception and production across phonological processes, language familiarities, and sensory modalities

Previous studies on the relation between speech perception and production took advantage of the great variability that L2 speakers show in these skills. However, these studies have not consistently found such a relationship (Flege et al., 1999; Kartushina & Frauenfelder, 2014; Levy & Law, 2010; Peperkamp & Bouchon, 2011; Rallo Fabra & Romero, 2012). In the third study (Chapter 4) we investigated the relationship of speech perception and production but in contrast with previous studies, we considered speech perception as a broad ability that involves several phonological processes, listener’s languages, and sensory modalities. To this end, we measured speech perception and production abilities in a large sample of early and highly proficient Spanish (L1)-Catalan (L2) bilinguals in a variety of linguistic and non-linguistic tasks. We measured speech perception skills across phonological processes (sub-lexical and lexical), listeners’ language familiarities (native, L2, and non-native) and sensory modalities (auditory and visual). Different phonological processes were tapped by a categorization, gating and lexical decision task in the L2 (Sebastián-Gallés & Baus, 2005). Native and non-native phoneme perception abilities were assessed by the MMN (Díaz et al., 2008) and visual speech processing ability was tested behaviorally and electrophysiological by a cross-modal prediction task (Sánchez-García et al., 2011, 2013). L2 speech production ability was assessed by a picture-naming task involving lexical processing for several vowels and consonants (Sebastián-Gallés & Baus, 2005). In addition to the linguistic measures, we measured individual differences in non-linguistic abilities such as the MMN elicited by tone discrimination (Díaz et al., 2008) and sensory-motor skills with a computerized drawing task. To see the complex interactions between the different measures, we run factor analysis.

Results of the factor analysis revealed four factors, yet only the first factor showed evidence for a tight relationship between speech perception and production which held across phonological processes (i.e., sub-lexical and lexical) and languages (native and

L2). This factor grouped together all L2 perception skills, L2 production skills, and the discrimination sensitivity (MMN response) for the native phoneme changes. Interestingly, highest factor loadings were seen for the speech production measures, suggesting that the prime process in the factor was speech production. This interpretation favors motor theories of speech perception which propose that the basic unit of speech perception is the articulatory gesture (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010).

The first factor also revealed that the type of phonological processes and the listeners' language familiarity did not seem to have a major impact on the strength of the relationship between speech perception and production. Although factor loadings were slightly higher for tasks involving words (lexical decision task, gating task, and picture naming task) as compared to isolated phonemes (categorization task and MMN to native phoneme), all tasks were included in the factor. Hence, the inconsistent results of previous studies investigating the correlation between L2 speech perception and production abilities cannot be accounted by the type of phonological process compared. Furthermore, the factor load for the native speech perception ability was lower than for the L2 speech perception and production measures. The relationship between native and L2 abilities in factor one is in line with previous studies that showed that good and poor perceivers of a difficult L2 phoneme contrast differed in their phoneme discrimination abilities of native and unknown phonemes, as measured by the MMN (Díaz et al., 2008; Díaz, Mitterer, Broersma, Escera, & Sebastián-Gallés, in press). The authors interpreted the finding as evidence for a speech specific origin of L2 perception skills. The results of the current study extended these previous results to speech production. Still, in the present study we found that the native MMN was incorporated in the factor with the lowest factor load and the MMN to the non-native phoneme was not present in the factor. These results suggest that the strength of the relation between perception and production is maximal within the same familiar language. The absence of the MMN to non-native phonemes is in line with Díaz et al. (2008)´s results of non-native phonemes eliciting the smallest MMN differences between L2 good and poor perceivers, though this was not submitted to statisticall tests.

Unexpectedly, auditory speech perception and production abilities did not relate to visual speech perception abilities. Rather, all measures for the audio-visual prediction task were grouped in a separate factor, the third factor. This pattern suggests that audio-visual integration of speech might primarily rely on general domain multisensory mechanisms. Furthermore, non-linguistic sensory motor skills did not relate to the speech perception and production tasks. All measures of the drawing task and the reaction time of the audio-visual task grouped together in a common factor, the fourth factor. The second factor integrated a variety of measures including the MMN to tones and phoneme changes, the N1 from the audio-visual prediction task, and L2 vowel production. We suggest that all measures are related to prediction processes.

Taken together, in this study we found a clear and strong relationship between speech perception and production skills. By means of factor analysis we were able to show that different measures for speech perception and production skills grouped together independently of the phonological and lexical processes involved and languages tested. Highest loadings in the factor corresponded to the speech production measures, suggesting that speech production plays an essential role in speech perception, as claimed by motor theories of speech perception (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010). Furthermore, the relationship between perception and production skills was relatively independent from other skills, showing few commonalities with audio-visual speech perception, general auditory skills, and general sensory-motor skills.

5.2. Implications for the speech perception and production literature

The results of this dissertation show the influence of speech production mechanisms in speech perception. By studying speech perception across language familiarities we were able to show that listening to non-native speech can activate speech production areas (Chapter 2). This result challenges acoustic theories of language learning that disregard any involvement of production in phoneme learning. Rather, they claim that speech production depends on speech perception because any phoneme should be accurately perceived before it can be produced (Flege, 1995; Kuhl & Iverson, 1995). Here we show that the production system attempts to model articulatory gestures for phonemes that are unknown. We interpret this finding as a suggestion that the production system may play an active role during phoneme learning. However, the present study was not designed to investigate the contribution of production mechanisms to phoneme learning and this conclusion should be taken with caution.

The contribution of speech production in perception reported in the first study (Chapter 2) is in line with the claim of the motor view that speech production is the core process supporting phoneme learning (Best, 1995; Best & Tyler, 2007; Fowler, 1986; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010). However the present results are not conclusive in regards to the nature of the relation between speech perception and production because they can be also accounted by an acoustic model. The Dual-Stream model of speech perception (Hickok & Poeppel, 2007) proposed cross-talks between speech perception and production when perception encounters difficulties in recognizing phonemes based on the acoustic cues, such as when the speech signal is degraded. Thus, the findings of the first study (Chapter 2) can be accommodated to both the acoustic and motor views of speech perception.

As presented in the introduction, the critical evidence for the motor view of speech perception would be finding an influence of speech production on perception even when the speech signal is not degraded and the perceived phonemes are known. This was tested in the second study (Chapter 3) with a somatosensory manipulation during passive listening of native and non-native phonemes. However, the results were inconclusive in regard to an influence of speech production in perception and, rather, revealed new insights regarding the effects of attention processes when a somatosensory manipulation is introduced in the experimental situation. Still, the factor analysis reported in the third study (Chapter 4) revealed a single factor for auditory speech perception and production abilities in which the higher load belonged to production measures. This finding suggests that production is not just a subsidiary process to perception, as proposed by the acoustic accounts, but rather a critical process to achieve accurate speech perception and is, hence, difficult to accommodate in acoustic proposals of speech perception. In addition, the factor analysis showed that L2 perception and production skills were related to discrimination sensitivity of native phonemes. This is the first time that native speech perception capabilities are shown to relate to L2 production. This novel finding implies the existence of a unique speech ability for both perception and production abilities that varies among individuals as the origin of individual differences in L2 learning.

To sum up, in this dissertation we report an involvement of speech production during perception and a strong relation of both abilities. Our results highlight the activation of the production system even when listening to non-native phonemes and show that individual variability in L2 speech perception and production are related to native speech perception abilities. Yet, the evidence here reported is not sufficient to reveal the nature of the relationship between speech perception and production. Future studies are needed to reveal whether speech perception and production share intended gestures as phoneme representations, as claimed by motor accounts, or rather speech perception recognizes phonemes based solely on acoustic features, as claimed by acoustic accounts.

5.3. Future lines of research

One question that remains unknown is whether changes in activations of motor areas to non-native phonemes, as the ones reported in the first study (Chapter 2), can be an index of phoneme learning. This question could be addressed by studying if the activations we found for initial exposure to non-native phonemes would change as the non-native phonemes are learned and become part of the listener’s repertoire. Based on our results, showing positive correlations between motor activations and the participants’ non-nativeness ratings of the phonemes, we would expect motor activations to decrease when L2 phoneme categories are established. Eventually, when a good command of the L2 phonemes is attained, we would expect motor activations to non-native phonemes to be comparable to those evoked by native phonemes.

An unexpected, though interesting result of the present research was the general attentional effect triggered by starting an experiment with a somatosensory manipulation, a spoon over the tongue or in the hand. This general attentional effect is probably due to monitoring the somatosensory manipulation during the perception task (taking care that the spoon does not fall down) and might have covered articulatory specific effects due to the spoon. To prevent general attentional effects in following studies, we suggest to manipulate the production system before the perception task, such as in previous studies employing rTMS over an articulatory muscle (Möttönen et al., 2013, 2014) or a somatosensory training before the study (Sato et al., 2011). Alternatively, a short somatosensory manipulation could be employed before a trial, as in a previous study (Ito et al., 2009).

Finally, in this dissertation we mentioned only shortly another debate in the speech perception literature: if speech is special or rather influenced by general domain abilities. Regarding the relationship between speech perception and production, a general domain ability that is closely related to speech is the learning of a musical instrument. Playing an instrument requires the acquisition of fine motor skills as well as the interaction between auditory and motor cortices (Zatorre, Chen, & Penhune, 2007). Previous studies showed that listening to music activates motor regions in musicians (Haueisen & Knösche, 2001) and that practicing a music piece results in increase in motor excitation when listening to the same music piece afterwards (D’Ausilio, Altenmüller, Olivetti Belardinelli, & Lotze, 2006). It remains to be studied how similar these motor interactions when listening to music are to speech perception and if they could influence each other.

6. REFERENCES

Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience, 23*, 4038–47.

Alho, J., Sato, M., Sams, M., Schwartz, J.-L., Tiitinen, H., & Jääskeläinen, I. P. (2012). Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. *NeuroImage, 60*, 1937–1946.

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *The European Journal of Neuroscience, 12*, 2570–2574.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*, 339–355.

Bartoli, E., D'Ausilio, A., Berry, J., Badino, L., Bever, T., & Fadiga, L. (2015). Listener-Speaker Perceived Distance Predicts the Degree of Motor Contribution to Speech Perception. *Cerebral Cortex, 25*, 281–288.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*, 233–252.

Besle, J., Bertrand, O., & Giard, M. H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research, 258*, 143–151.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204). Baltimore: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in* *second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing Company.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program].

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*, 189–221.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, *64*, 74–95.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, *22*, 1182–94.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*, 113–124.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593–596.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., … David, A. S. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*, *12*, 233–243.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the Graphemic Buffer in spelling: evidence from a case of acquired dysgraphia. *Cognition, 26,* 59–85.

Cheour, M., Alho, K., Ceponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., … Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, *29*, 217–226.

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA : Using the SPSS R-Menu v2 . 0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, *18,* 1–14.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*, 3668–3678.

D’Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*, 955–958.

D’Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, *118*, 9–14.

D’Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex, 48,* 882–887.

D’Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: priming and coarticulation effects. *Neuropsychologia*, *49*, 3670–3676.

D’Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.

Deguchi, C., Chobert, J., Brunellière, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. *Brain Research*, *1366*, 149–161.

Dell, G. S. (1986). A spreading-activation theory of retreival in sentence production. *Psychological Review*, *93,* 283–321.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*, 176–180.

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 16083–16088.

Díaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastián-Gallés, N. (2015). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*, 680–689.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., … Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*, 1883–1908.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress "deafness" in simultaneous French-Spanish bilinguals. *Cognition*, *114*, 266–275.

Erlbeck, H., Kübler, A., Kotchoubey, B., & Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Frontiers in Human Neuroscience*, *8*, 1–16.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*, 399–402.

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Baltimore: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–355). Berlin: Mouton de Gruyter.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*, 2973–2987.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52.

Gabay, Y., Schiff, R., & Vakil, E. (2012). Dissociation between the procedural learning of letter names and motor sequences in developmental dyslexia. *Neuropsychologia*, *50*, 2435–2441.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind*. New York, NY: W. W. Norton & Company.

Grabski, K., Schwartz, J. L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., … Sato, M. (2013). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, *26*, 384–408.

Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain Research*, *1515*, 55–65.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13*, 786–792.

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,* 360–378.

Hedenius, M., Persson, J., Alm, P. a., Ullman, M. T., Howard, J. H., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities*, *34*, 3924–3935.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron, 69,* 407–422.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*, 393–402.

Hillis, A. E., Rapp, B. C., & Caramazza, A. (1999). When a rose is a rose in speech but a tulip in writing. *Cortex*, *35*, 337–356.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 1245–1248.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5,* 1–17.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, *8*, 43–52.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297,* 846–848.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect.” In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-*

Language Research (pp. 121–154). Maryland: York Press.

Leung, M.-T., & Ciocca, V. (2011). The effects of tongue loading and auditory feedback on vowel production. *The Journal of the Acoustical Society of America, 129,* 316–325.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, Massachusetts: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America, 126,* 2670–2682.

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*, 1290–1305.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109,* 1431–1436.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253–257.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., … Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport, 11,* 1729–1733.

Majewski, W., Zalewski, J., & Hollien, H. (1979). Some remarks on different speaker identification techniques. In H. Hollien & P.

Hollien (Eds.), *Current Issues in the Phonetic Sciences* (pp. 829–835). Amsterdam: John Benjamins.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing Interactions Word Recognition and Lexical Access during in Continuous Speech. *Cognitive Psychology, 10,* 29–63.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18,* 1–86.

McClelland, J. L., Mirman, D., & Holt, L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10,* 363–369.

McGee, T. J., King, C., Tremblay, K., Nicol, T. G., Cunningham, J., & Kraus, N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, *38,* 653–658.

McGurk, H., & MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature*, *264*, 746–748.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*, 1692–6.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of Linguistic Experience: The discimination of [r] and [l] by native Speakers of Japanes and English. *Perception & Psychophysics*, *18,* 331–340.

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, *23,* 1190–1197.

Möttönen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *The Journal of Neuroscience, 34*, 4064–4069.

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience, 29*, 9819–9825.

Möttönen, R., & Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology, 26*, 1103–1118.

Murakami, T., Restle, J., & Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia*, *49*, 2045–54.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*, 2544–2590.

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, *115*, 140–144.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, *24*, 283–288.

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 20470–20475.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*, 4–12.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, *92*, 26–32.

Norris, D. (1994). Shortlist - a Connectionist Model of Continuous Speech Recognition. *Cognition*, *52*, 189–234.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage, 54,* 2437–45.

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology*, *82*, 219–226.

Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., … Huotilainen, M. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *International Journal of Psychophysiology*, *87*, 103–110.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*, B9–B17.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. a, De Giovanni, U., … Fazio, F. (2003). A functional-anatomical model for lipreading. *Journal of Neurophysiology*, *90*, 2005–2013.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 161–164).

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, *26*, 490–498.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception, *106*, 10598–10602.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature*

Reviews. Neuroscience, *11*, 351–360.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7865–7870.

R. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*, 491–508.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*, 212–218.

Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*, 712–722.

Rizzolatti, G., & Arbib, M. a. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188–194.

Rogers, J. C., Möttönen, R., Boyles, R., & Watkins, K. E. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology*, *5*, 754.

Ronquest, R. E., Levi, S. V, & Pisoni, D. B. (2010). Language identification from visual-only speech signals. *Attention, Perception & Psychophysics*, *72*, 1601–1613.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*, 1147–1153.

Roy, A. C., Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *Journal of Physiology*, *102*, 101–5.

Sánchez-García, C., Alsius, A., Enns, J. T., & Soto-Faraco, S. (2011). Cross-modal prediction in speech perception. *PLoS ONE*, *6*, 1–8.

Sánchez-García, C., Enns, J. T., & Soto-Faraco, S. (2013). Cross-modal prediction in speech depends on prior linguistic experience. *Experimental Brain Research, 225*, 499–511.

Sato, M., Buccino, G., Gentilucci, M., & Cattaneo, L. (2010). On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Communication, 52*, 533–541.

Sato, M., Grabski, K., Glenberg, A. M., Brisebois, A., Basirat, A., Ménard, L., & Cattaneo, L. (2011). Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. *Cortex*, *47*, 1001–1003.

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*, 1–7.

Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*, 336–354.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action-candidate roles for the motor cortex in speech perception. *Nature Reviews. Neuroscience, 10*, 295–302.

Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-first century Psycholinguistics: Four cornerstones* (pp. 279–292). New York: Erlbaum.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111–123.

Soto-Faraco, S., Navarra, J., Weikum, W., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics, 69*, 218–231.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*, 1964–1973.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naive American listeners. *The Journal of the Acoustical Society of America*, *126*, 1461–1476.

Sumby, W. H. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*, 212.

Sundara, M., Namasivayam, A. K., & Chen, R. (2001). Observation-execution matching system for speech: a magnetic stimulation study. *Neuroreport*, *12*, 1341–1344.

Sussman, E., Ritter, W., & Vaughan, H. G. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, *789*, 130–138.

Sussman, E. S. (2013). Attention Matters: Pitch vs. Pattern Processing in Adolescence. *Frontiers in Psychology*, *4*, 1–8.

Sussman, E., Winkler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, *40*, 430–435.

Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., & Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain and Language*, *126*, 1–7.

Szymanski, M. D., Yund, E. W., & Woods, D. L. (1999). Phonemes, intensity and attention: differential effects on the mismatch negativity (MMN). *The Journal of the Acoustical Society of America*, *106*, 3492–3505.

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, a. D. (2009). Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience, 30,* 1636–1642.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270.

Vicari, S., Finzi, a, Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery, and Psychiatry, 76,* 1392–1397.

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22,* 1583–1596.

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience, 32,* 3665–3678.

Wassenhove, V. Van, Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech Neurophysiological Basis of Multisensory Integration. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 1181–1186.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41,* 989–994.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1–25.

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. (2007). Visual Language Discrimination in Infancy. *Science, 316,* 1159.

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience, 7,* 1–8.

Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the

First Year of Life, *7*, 49–63.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage, 33*, 316–325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*, 701–702.

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*, 132–143.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, a, … Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638–642.

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*, 30–42.

Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., & Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, *35*, 283–92.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8,* 547–558.