Before phonemes: Infants start building the native phoneme repertoire

Konstantina Eirini Zacharaki

TESI DOCTORAL UPF / 2021

Directora de la tesi

Dra. Núria Sebastián-Gallés,

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

Universitat Pompeu Fabra Barcelona

To my family,

Acknowledgments

I would like to take this opportunity to express my gratitude to Nuria. I would like to thank her for everything she has taught me in the last four (five if we count the master) years. It has been a true honor and privilege to do my PhD under her supervision. Thank you for entrusting me with these projects. It has been a dream come true.

I also feel the need to thank two of my professors from my undergraduate studies at the University of Athens. Professor Hatzidaki for introducing me into the academic world and making me read academic articles for the first time (two of those were the classic Bosch & Sebastian-Galles, 1997, 2001). I would also like to thank professor Georgiafentis for transmitting to me his love for language and Chomsky.

Gracias a todas las personas que han pasado por el grupo SAP. Ya desde el primer momento, en el máster, me sentí muy bien acogida y que formaba parte de una familia. Gracias a Miguel por hablar con Núria sobre mi futuro en Bilbao. Gracias al equipo de las conferencias: Marc, Jesús, Chiara, Gonzalo, Camille y Joan. Hemos aprendido mucho juntos y siempre me he sentido acompañada, especialmente cuando dejábamos nuestra burbuja de la uni y pasamos a un mundo bastante competitivo. Gracias a Cami, Sànder y Paola por compartir las clases de bachata y salsa en la Pompeu. Gracias a Carlos por enseñarme que la academia se puede combinar con el baile y por enseñarme que la vida se vive mejor cuando te rodeas de gente especial y llena de amor. Gracias a mis compis de despacho: Marc, Manel, David, Ludo, Ane, Gonzalo, por el interés que teníamos mutuamente, los diálogos filosóficos que teníamos sin llegar a ningún lado y por ser el mejor despacho con la mayor variedad de cafés y tes. Gracias a Indre, Devin y Farhad por comer con nosotros cada día y ser nuestros amigos de otros grupos de investigación. Gracias a Cami por los consejos, aun no estando en el grupo ahora, y por invitarnos a su casa en Poitiers. Gracias sobre todo a Chiara, Ege y Gonzalo, aka a 'los guys' por haber compartido cada momento del doctorado juntos y formar nuestro grupito de apoyo moral. Gracias por todas las experiencias, cenas y botellas de vino que hemos compartido. Ahora parece que me estoy doctorando al nivel de las abuelas (vida postdoctorado). A nivel profesional, gracias a Chiara por aconsejarme cuando yo no sabía que tenía que hacer o como organizarme. Has sido como una hermana mayor en el trabajo. Gracias a Gonz por transmitirme su amor hacia R y por enseñarme que aplicar la estadística no tiene por qué ser tan difícil. Gracias a Ege por compartir conmigo sus conocimientos de EEG y aguantarme cuando trabajábamos juntas. Gracias a Ed por aguantarnos (Ege, Gonzalo y a mí) porque tiene que ser muy complicado ser amigo de 3 estudiantes de doctorado. También gracias a Ed por haber corregido el texto de esta tesis. Gracias a Carlota por los momentos que hemos compartido tanto en la uni, como en nuestro piso (y compartiendo a nuestro gato). Por último, me gustaría agradecer a Ane. He sido muy afortunada de haberte conocido. Gracias por tu apoyo y amor incondicional, gracias por enseñarme a percibir las cosas con otra perspectiva y por ser una fuente de luz y de amor pasase lo que pasase.

Me gustaría agradecer también a todos los miembros del CBC teniendo en cuenta que cada uno me ha aportado algo diferente en esta experiencia. Tengo que dar las gracias a nuestros técnicos, Xavi y Silvia. Sin vuestra ayuda mi trabajo hubiera sido mil veces más difícil. Gracias a Katia y Cris, las dos Babylab managers, que se encargaron del laboratorio durante mi doctorado. No me puedo olvidar de dar las gracias a Pam, Cristina, Bea y Florencia por su ayuda a nivel administrativo. También aprovecho para agradecer a Cinta por todo el trabajo que realizó en el proyecto de discriminación de lenguas. Gracias a todas las asistentes que han trabajado estos años en el Babylab, Cinta, Alba, María, Aina y Andrea. Por último, agradecer a todas las familias y los bebés que participaron en los estudios. Fue un placer conoceros y aprender y jugar con vosotros.

Πάνω από όλα θα ήθελα να ευχαριστήσω την οικογένεια μου. Ευχαριστώ μαμά και μπαμπά για την ανιδιοτελή αγάπη και στήριξη σας πριν καν γεννηθώ. Ευχαριστώ που μου με μάθατε να πιστεύω στον εαυτό μου και ότι μπορώ να θέτω υψηλούς στόχους. Ευχαριστώ για τις αρχές που μου δώσατε, για όλες τις θυσίες που κάνατε ώστε να έχω όσο το δυνατόν καλύτερη εκπαίδευση αλλά και για την εμπιστοσύνη που μου δείξατε στο να κάνω τις δικές μου επιλογές. Θα ήθελα να πω ένα μεγάλο ευχαριστώ στη μαμά μου που με στήριξε στο να μάθω ισπανικά ενώ σπούδαζα. Ποιος ήξερε ότι τελικά θα μείνω στην Ισπανία για 5 χρόνια; Θα ήθελα να

ευχαριστήσω και την αδερφή μου για την στήριξη και την αγάπη της καθ' όλη την διάρκεια της ζωής μας. Νομίζω εκείνη είναι πιο περήφανη από εμένα που έκανα το διδακτορικό. Θα ήθελα να ευχαριστήσω και την Κοραλία για την αγάπη της και την πραγματική φιλία της. Τέλος θα ήθελα να ευχαριστήσω και το ζευγάρι μου, τον Δημήτρη, που δεν με ανάγκασε να διαλέξω ανάμεσα από τους ακαδημαϊκούς μου στόχους και την προσωπική μου ευτυχία.

Abstract

Infants start their lives with a universal ability to perceive speech and during the first months of life they attune to the language(s) they are exposed to in their environment, i.e. perceptual narrowing. Research has focused on infants' capacities to discriminate native and non-native speech contrasts as a sign of this tuning, starting at 6 months of age for vowels (Kuhl et al., 1992; Polka & Werker, 1994). We investigated whether infants before the first signs of perceptual narrowing have some segmental information in place. To do so we ran a series of experiments on the abilities of infants to discriminate languages that differ in their vowel distribution. We also tested infants' preference to lists of nonwords that abide to the vowel distribution of their native language or not. We found that infants succeeded in both tasks suggesting that infants have in place an early representation of the native vowel space. Therefore, we provide compelling evidence that phonetic knowledge emerges earlier than proposed before.

Resumen

Los infantes comienzan su vida con una habilidad universal para percibir el habla y durante los primeros meses de vida se especializan en la lengua o lenguas que escuchan en su entorno (estrechamiento perceptual). Las investigaciones previas se han centrado en la capacidad de los bebés para discriminar contrastes del habla nativos y no nativos como evidencia de este estrechamiento, a partir de los 6 meses de edad para las vocales (Kuhl et al., 1992; Polka & Werker, 1994). En esta tesis investigamos si los bebés poseen conocimiento segmental antes de los primeros signos del estrechamiento perceptual. Para ello, llevamos a cabo una serie de experimentos sobre su capacidad de discriminar entre lenguas con distribuciones vocálicas diferentes. También investigamos la preferencia de los bebés por unas listas de palabras inventadas que reflejan o no la distribución vocálica de su idioma nativo. Hallamos que los bebés realizaron con éxito ambas tareas, lo que sugiere que tienen una representación temprana del espacio vocal nativo. Por lo tanto, proporcionamos evidencia convincente de que el conocimiento fonético surge antes de lo propuesto anteriormente.

Σύνοψη

Τα μωρά ξεκινούν την ζωή τους με μία γενική ικανότητα να αντιλαμβάνονται την γλώσσα και κατά την διάρκεια των πρώτων μηνών της ζωής τους προσαρμόζουν την αντίληψη τους στην γλώσσα που ακούν στο περιβάλλον τους, η λεγόμενη αντιληπτική στένωση. Οι προηγούμενες έρευνες έχουν επικεντρωθεί στις ικανότητες των βρεφών να διακρίνουν ήχους που προέρχονται από την μητρική τους γλώσσα ή από μία ξένη γλώσσα ως ένδειξη αυτής της προσαρμογής, ξεκινώντας από την ηλικία των 6 μηνών για τα φωνήεντα (Kuhl et al., 1992; Polka & Werker, 1994). Εμείς ερευνήσαμε αν τα μωρά έχουν φωνήματα. Για να το υλοποιήσουμε, πραγματοποιήσαμε μια σειρά πειραμάτων σχετικά με τις ικανότητες των βρεφών να διακρίνουν γλώσσες που διαφέρουν ως προς την κατανομή των φωνηέντων τους. Επίσης, εξετάσαμε την προτίμηση των μωρών χρησιμοποιώντας λίστες λέξεων που είτε συμμορφώνονται με την κατανομή των φωνηέντων της μητρικής τους γλώσσας είτε όχι. Βρήκαμε ότι τα βρέφη ολοκλήρωσαν με επιτυχία και τα δύο πειράματα, υποδηλώνοντας ότι έχουν μια πρώιμη ιδέα των φωνηέντων της μητρικής γλώσσας. Επομένως, αποδεικνύουμε ότι η φωνητική γνώση αναδύεται νωρίτερα από ότι είχε προταθεί.

Preface

Hi, I am Konstantina, and I am a language aficionada. Since I was a teenager, I was amazed by the impact that language has on perception and on others. The way we say something, our choice of words, can really affect how others interpret the meaning we want to portray. There is a Greek saying that goes, 'the tongue does not have bones, but it can break them' (Η γλώσσα κόκκαλα δεν έχει και κόκκαλα τσακίζει). That is how powerful language can be. Language is also one of the main features that distinguishes us from other animals. The origins of linguistic knowledge have been attributed to different sources, giving birth to the conundrum of 'nature versus nurture'. One common factor across these theories is the need to understand how humans acquire this powerful tool, language, in the presence of limited experience and no explicit supervision, i.e., Plato's problem. Infants are born with general abilities to perceive speech and during the first years of life they accomplish learning the specificities of the complex prosodic, phonological, grammatical, and semantic systems of their native language. My goal in this dissertation has been to find the intersection of general and native knowledge on the segmental information.

Infants during the first months of life can distinguish sounds regardless of whether they belong to the native phoneme repertoire or not. Infants gradually attune to the native language, a phenomenon known as *perceptual narrowing*. By 6-8 months vowel perception is modulated by the native language and by 10-12 months they can no longer discriminate non-native consonantal contrasts (Kuhl et al., 1992; Werker & Tees, 1984). Although, this line of research has provided a vital insight of early perceptual abilities, it might have underestimated the emergence of such knowledge. The proposed mechanism behind the establishment of phonemes is *distributional learning*. Infants accumulate information on the frequency of sounds from their exposure to language that eventually allows them to group them into categories. It has been shown that the mechanism behind phonetic learning is active very early on for vowels, i.e. at 2-3 months of age (Wanrooij et al., 2014). Taking into consideration that the first signs of perceptual narrowing for vowels occur almost 3 months later, it raises the question about the precursors of phonetic knowledge.

The aim of this dissertation is to investigate if infants have protosegmental information at their disposal. To do so, I ran a series of experiments. First, I tested 4.5-month-old infants on their capacity to discriminate languages based on vowel distribution. One of the pairs of languages had noticeable differences, Eastern Catalan and Western Catalan, while the other pair did not, Western Catalan and Spanish. Second, I investigated whether infants would show a preference for lists of nonwords that reflect the vowel distribution of their native language. Additionally, speech perception in adulthood was also explored concerning intonational and segmental information.

Unfortunately, during the course of my PhD, a global pandemic hit. The worldwide phenomenon had an immense impact on our lives and work. As we were collecting data, almost halfway, on the experiment on vowel preference with infants (reported in Chapter 4), the pandemic of Covid-19 reached Spain. Citizens of Catalonia were confined in their homes for 98 consecutive days, which took a toll on our research, among other things. Apart from not being able to collect data in the laboratory, we were also excluded from recruiting participants from two major clinics in Barcelona (20 months later we still are), which were our main sources of participants. Luckily, we were able to resume testing in July 2020, but data collection was painfully slow. A new health safety protocol was in place that made the duration of the appointments longer to ensure the sanitization of the laboratory between appointments. A fear of contamination was also present. Some families were skeptical of coming to our premises and to collaborate. For myself, it was also no longer mundane to take the metro to go to work and come in contact with families. Thankfully, we were able to complete our sample size of the aforementioned experiment. However, we were unable to run the additional experiments we had thought would be included in the present dissertation. Given the circumstances, we had to adapt to the new status quo to collect enough data and to be able to submit the thesis. Thanks to my supervisor and the help of the technicians working in the Center for Brain and Cognition, we were able to design and run two online experiments with adults in a new platform (Chapter 3 and Chapter 5). Although the pandemic has affected our work, we believe in the merit of the research reported in this dissertation. Personally, I feel lucky that Covid-19 has only affected the research plan.

Table of contents

1. Introduction

Language acquisition is the process by which infants learn their native language(s) that will allow them to understand and produce it. This happens progressively and specific milestones have been discovered throughout the first years of life. The topic of this thesis is to investigate what happens between two periods extensively investigated in infant research: the neonatal period (0-2 months of age) and the second half of the first year of life. It is surprising that we know so much about the abilities of newborns, and on language development after 6 months of age, whereas we know so little about the period from 2 to 6 months. There is a general consensus that newborns have universal capacities to perceive speech, while at 6 months, speech perception is language-specific (figure 1.1.). However, it is unrealistic to believe that nothing happens in such a protracted period of time. The goal of this dissertation is to focus right before the first evidence of language-specific speech perception. The introduction is organized following a methodological thread: we will first review research on language discrimination and then review research on the establishment of the phonetic repertoire. The first section of this thesis focuses on the relevant theoretical and methodological background on infant research, and it also covers, without so much depth, research on adults.

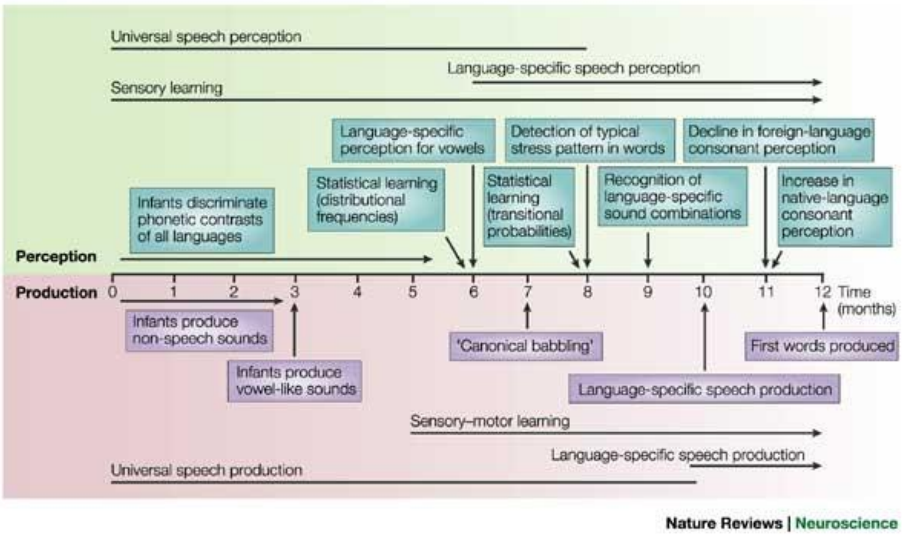


Figure 1.1. Reproduced from Kuhl (2004). The figure represents the developmental changes in speech perception and production during the first year of life.

1.1. Language discrimination in the first months of life

1.1.1. Rhythm - Theoretical background

Infants are raised in linguistically rich environments, often exposed to more than one language. Therefore, it is reasonable to assume that infants are equipped with mechanisms that will quickly allow them to separate languages, regardless of the environment they actually grow up in. Infants can discriminate between languages that differ prosodically from birth and there is some evidence that it can happen while in utero too (Mehler et al., 1988; Minai et al., 2017). The ability to discriminate between languages that sound more similar emerges a few months later around the fourth month of life (Bosch & Sebastian-Galles, 1997, 2001; Nazzi et al., 2000). Previous research has proposed that the initial type of information infants have at their disposal is rhythm.

Rhythm stems from the Ancient Greek word 'ρυθμός' and it means flow. This recurrent flow was noticed first by Lloyd James (1940) who described languages like English as 'morse-code' and languages like French as 'machine-gun'. These two categories were formalized as 'stress-timed' and 'syllable-timed' languages respectively (Pike, 1945). The difference between these categories was said to correspond to the isochrony of the temporal unit used in each category, between interstress intervals or between syllables (Abercrombie, 1967). A third category named 'mora-timed' was added to describe languages like Japanese (Ladefoged, 1975). The isochrony hypothesis has been disproven (Roach, 1982), which gave rise to a new proposal according to which languages are organized based on their phonological properties (Dasher & Bolinger, 1982). Dauer (1983) proposed that syllable complexity and vowel reduction are key distinctive phonological characteristics of syllable and stress timed languages. She also proposed that languages are organized in a continuum where the two endpoints are syllable and stress timed languages. The placement of a particular language on the spectrum is a result of rating its features.

An extension of this phonological account was provided by Mehler et al. (1996) who put forward the framework of Time and Intensity Grid REpresentation (TIGRE). The authors' goal was to provide an explanation of how infants discriminate between languages and subsequently how they bootstrap relevant information of their native language. They proposed that infants process speech by paying attention to the vocalic nuclei and that the intervocalic distance encapsulates the periodicity of the speech signal. The implementation of this framework was done a few years later when researchers analyzed sentences in eight different languages (English, Dutch, Polish, French, Spanish, Italian, Catalan, Japanese) trying to determine the 'correlates of rhythm'. To do so, they transcribed the vocalic and consonantal intervals in the sentences, which yielded three variables: i) the proportion of vocalic intervals (%V), ii) the standard deviation of the duration of vocalic intervals (ΔV) and iii) the standard deviation of consonantal intervals (ΔC). The results of this research indicated that the combination of %V and ΔC categorized languages better in terms of rhythm (figure 1.2.) (Ramus et al., 1999; for more recent accounts also see Langus et al., 2017; Nespor et al., 2011). This ternary distinction has been a long-standing conundrum. Other metrics and approaches have also been proposed to capture rhythm (Frota & Vigário, 2001; Grabe & Low, 2002; Nolan & Jeon, 2014; Tilsen & Arvaniti, 2013; Wagner & Dellwo, 2004; White & Mattys, 2007; for a meta-analysis on rhythm and metrics see Gasparini et al., 2021) but they have not been tested systematically with infants. The usefulness of metrics has also been questioned (Arvaniti, 2009, 2012). Infant studies provide support to the existence of the three rhythmic categories.

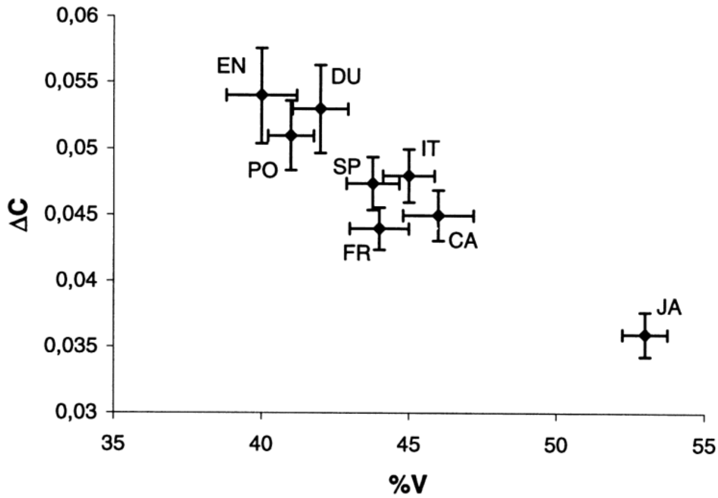


Figure 1.2. Reproduced from Ramus et al. (1999). Stress timed languages are grouped on the top-left, syllable-timed languages in the middle and mora-timed at the bottom-right.

1.1.2. Rhythm – Experimental evidence

Early investigation on infants' ability to discriminate languages has used different kinds of stimuli: natural, low-pass filtered and resynthesized speech being the most common. Natural speech contains both segmental and prosodic information, whereas in low-pass filtered speech most segmental information is eliminated. A specific type of speech resynthesis was used by Ramus & Mehler (1999) that can degrade the speech signal to different degrees, leaving behind different kinds of information. The first resynthesis method known as 'saltanaj' consists in transforming all consonants to the most common ones across languages for each category based on manner of articulation. Therefore, all fricatives are replaced by the phoneme /s/, all stop consonants with /t/, all liquids with /l/, all glides with /j/, and all nasals with /n/. All vowels are transformed by /a/ as it is the most common vowel universally. This transformation preserves broad phonotactics, rhythm and intonation. The second type of resynthesis is the 'sasasa', where vowels are replaced by 'a' and all consonants by 's'. This transformation conveys rhythm and intonation. The third type is called 'aaaa'. All phonemes are swapped with an 'a' and only intonation is preserved. The last resynthesis technique is the 'flat sasasa' which is the same as ‘sasasa’ only that it does not contain pitch modulations (for a summary of the techniques see table 1).

Neonates have been tested with different pairs of languages, either belonging to the same rhythmic class such as English and Dutch (stress-timed) or Spanish and Catalan (syllable-timed)[1], or languages from different rhythmic classes such French (syllable-timed) and Russian (stress-timed) (Mehler et al., 1988; Nazzi et al., 1998). The findings indicated that newborns could only discriminate between languages that belong to different rhythmic classes but failed to do so for languages that belong to the same rhythmic class. Their ability for between class discrimination was not hindered even when they were presented mixed sentences from two languages per rhythmic class (English + Dutch vs. Italian + Spanish). No discrimination was observed when the mixing of sentences belonged to two distinct rhythmic classes (English + Italian vs. Dutch + Spanish). The fact that the stimuli in these experiments were either natural or low-pass filtered suggested that neonates probably used rhythm to perform the task (see also Byers-Heinlein et al., (2010) for neonatal language discrimination of English, stress-timed, and Tagalog, mora-timed). Other work on neonatal language discrimination has found language discrimination with natural sentences, regardless of number of speakers (Mehler et al., 1988; Moon et al., 1993). Overall, previous work on language discrimination with newborns has provided strong evidence that infants use rhythm.

Table 1**.** Summary of resynthesis techniques used in language discrimination and the information that each one preserves.

Resynthesis techniques	**Information preserved**
Saltanaj	intonation, rhythm and phonotactics
Sasasa	rhythm and intonation
Aaaaa	only intonation
Flat sasasa	only rhythm

Computational modelling using i-vector systems has also been used to simulate this broad ability to discriminate languages. A Gaussian

[1] Unpublished data mentioned in Ramus (2002).

Mixture Model that was trained briefly on French, similar to the neonate's experience, was only able to distinguish languages that are rhythmically dissimilar (French-English), while failed to do so for languages that belong to the same rhythm class (Spanish-Catalan) (Carbajal, 2018; Carbajal et al., 2016).

The ability to discriminate between languages that differ rhythmically is not human specific. Previous research has shown that mammals, cotton-top tamarin monkeys and Long-Evans rats can discriminate forward Dutch from Japanese speech (Ramus et al., 2000; Toro et al., 2005). Additional evidence on tamarin monkeys shows that they, similarly to newborns, cannot discriminate languages from the same within class, English and Dutch (Tincoff et al., 2005). The results from non-human animals suggest that the ability to detect a signal that is different in terms of rhythm and melody is ancestral, employing general-domain acoustic abilities and that this capacity in infants has not evolved exclusively for language.

By two months, infants appear to have accumulated some additional experience with their native language, which affects their discrimination capacities. Discrimination between rhythmic categories is preserved and rhythm appears to be enough to do so, given that infants attended differently to sentences in English and French when they were either natural or low-pass filtered (Dehaene-Lambertz & Houston, 1998). However, infants cannot distinguish two languages that belong to different rhythmic classes when both are unfamiliar in terms of rhythmic properties. Christophe & Morton (1998) reported that English 2-month-olds could discriminate their native language from Japanese but failed in discriminating French from Japanese. An interesting pattern of results arose when they were presented with sentences in English and Dutch. Some of the infants were able to discriminate the two and others did not, yielding only a trend in discrimination. A similar trend was found when they were tested with Dutch and Japanese sentences, indicating that some participants might be treating Dutch as if it was their native language. The authors propose that they have tapped on a transitional period during which infants are adding information in the growing body of suprasegmental knowledge of their native language. Partial within class discrimination has also been observed with Spanish and/or Basque 3.5 month-old infants (Molnar et al., 2014).

Mixed results were also found in dialect perception. Dialects in most cases belong to the same rhythmic class[2] and have very similar phonetic repertoires, which makes them a great candidate for within class discrimination tasks. Australian 3-month-old infants showed a preference for their native dialect over American English in a central-fixation paradigm (Kitamura et al., 2013), while French infants of the same age showed no discrimination, when they were presented with Quebecois and Parisian French audio-visual recordings using fNIRS (Cristia, Minagawa-kawai, et al., 2014). Infants appear to experience a transitional period around the 2nd to 3rd month of life in which they maintain their ability to discriminate between rhythmic classes but have some limited ability to discriminate within the same category.

By 5 months of age, infants have gathered enough information on their native language. They can perform both within and between class discrimination, if one of the two languages is their native one. Given that some of the results reported in the following paragraph stem from the same rhythmic class, it stands to logic that even if participants have access to rhythmic cues, segmental and/or intonational cues, rhythmic ones will not be equally informative. Therefore, if discrimination is observed it would be due to other factors than rhythm. We report such findings in the present section because previous researchers have framed their research within the rhythmic hypothesis framework. Four-and-a-half-month-old infants orient faster to utterances in their native language (Spanish or Catalan), i.e. in comparison to utterances in a foreign language, i.e. English. This effect was observed even when the two languages, native and foreign were prosodically similar, Catalan and Spanish, using either natural or low-pass filtered sentences. This indicates that rhythmic information was enough to enable faster recognition of the native language. Subsequent work showed that both monolinguals and bilinguals could discriminate between Spanish and Catalan despite their rhythmic similarity (Bosch & Sebastian-Galles, 2001). Nazzi et al. (2000) put forward three different

---

[2] For instance, European Portuguese is stressed-timed and Brazilian Portuguese is syllable-timed (Frota et al., 2002).

hypotheses on infants’ language discrimination abilities during the first months of life. The first one was the *rhythm hypothesis* which entailed infants’ sensitivity to the native rhythmic class. According to the authors, infants could use rhythmic features to classify utterances in the rhythmic class (between class discrimination). The focus on the native rhythmic class would eventually allow infants to perform finer analyses of the rhythmic characteristics of languages. This hypothesis cannot by itself explain within class discrimination. The second hypothesis was the *native language acquisition*, which posited that infants’ knowledge of the native language evolves during the first months of life and once they have enough information of its rhythmic properties, they would be able to discriminate their native language from others within the same class. Lastly, the *maturational hypothesis* postulated that infants can discriminate between languages that are acoustically different within the native rhythmic class or a foreign one. All three hypotheses provide alternative developmental trajectories. Nevertheless, the authors failed to specify the rhythmic properties that could allow infants to perform within class discrimination (*rhythmic and native language hypotheses*). They also failed to mention if other parts of speech, i.e. intonation or segmental cues, would play a role. Nazzi and coworkers performed a series of experiments using different languages in order to test these hypotheses. They found that 5-month-old American infants can discriminate between rhythmic classes, (Italian vs. Japanese), but also within class (British English vs. Dutch). The authors also showed that within class discrimination is gated by familiarity to the languages tested. Participants failed to discriminate between languages coming from the same rhythmic class regardless of whether it was in the native one (German vs. Dutch, stress-timed) or in an unfamiliar one (Italian vs. Spanish, syllable-timed) because none of them was familiar. Though they were able to discriminate between their native dialect and British English, a fine distinction involving differences both at the segmental and intonational level. This pattern of results provided support to the *native language acquisition* hypothesis. They concluded that to describe language discrimination from birth to 5 months of age, the best account is a combination of the *native language* that *the rhythmic* hypotheses. The information that infants use for within class discrimination was debated in the discussion. Although the authors did not mention other phonological information other than rhythm in their hypotheses, they proposed that infants might have used prosodic (rhythm and intonation) or phonetic-phonotactic information to discriminate. Butler et al. (2011) provided converging evidence on the relevance of familiarity showing that 5-month-old infants learning British English could discriminate their native dialect from an unfamiliar Welsh English dialect, while they failed in distinguishing sentences from two unfamiliar English dialects.

Language discrimination at this age has also been studied using electrophysiological techniques. Peña et al., (2010) investigated how 3- and 6-month-old infants[3] processed sentences in their native language (Spanish), in a rhythmically similar one (Italian) and a rhythmically dissimilar one (Japanese). The authors looked at differences in gamma-power as a neural marker of perceptually binding elements that are presented separately. At 3 months of age, infants had greater gamma-power for the Spanish and the Italian sentences than for the Japanese ones. At 6 months, gamma-modulation was specific to the native language of the participants, i.e. it was significantly greater for Spanish than the other languages. They also found a shorter latency in the early ERP component P200 for the older group of participants. This was interpreted as a maturational sign of auditory processing. This pattern of results showing an initial rhythmic sensitivity that evolves to a native language specific one converges with the results found with behavioral paradigms. The authors suggested that infants may use segmental information and syllabic structures to succeed in discrimination (see Nacar Garcia et al., 2018 for similar results).

Adults have long established their native phonetic repertoire and can use such knowledge to judge whether an excerpt is foreign or not in a matter of seconds (Muthusamy et al., 1994). It has also been investigated whether adults can make use of rhythmic cues to discriminate between languages or not yielding mixed results. Ramus & Mehler (1999) tested the ability of French adults to discriminate between English (stress-timed language) and Japanese (mora-timed language) using the resynthesis techniques described earlier (see table 1). Discrimination scores showed that participants could distinguish between the two languages significantly above

[3] They tested both full-term and preterm infants. Here we focus on the full-term ones.

chance in all conditions ('saltanaj', 'sasasa', 'flat sasasa') but not in the condition that preserves only intonational cues ('aaaa'). Further evidence indicated that discrimination is driven by rhythm in adulthood. French adults were able to discriminate languages that belong to different rhythmic classes (English-Spanish, Catalan-English) when they only had access to rhythmic cues but failed to perform within class discrimination (English-Dutch, Catalan-Spanish)(Ramus et al., 2003).

1.1.3. Intonation – Theoretical background

From a Linguistic point of view, Intonation has been studied extensively but there is no consensus on what exactly it comprises. There are two different definitions. According to the broad definition, the terms of intonation and prosody can be used interchangeably, meaning that it includes both rhythm and melody of speech (Allen, 1971). The narrow definition of speech refers to the linguistic use of pitch, i.e. the fundamental frequency (F0) of voiced sounds (for more information on the two definitions see Hirst & Di Cristo, 1998; Levis, 1999). Our use of the term intonation identifies more with the narrow definition which was described nicely by Ladd (2008, p.4) as 'the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way'.

Intonation is a complex part of speech that has been described both from the phonological and the acoustic standpoints. The phonological account has been inextricably linked to cross-language research (Bolinger, 1982; Grover et al., 1987; Hirst & Di Cristo, 1998; Jun, 2010). One common aspect across languages is intonation's functions. Firstly, intonation is used for lexical and morphological marking utilizing pitch, which divides languages into tone languages and intonation languages. Tone languages such as Chinese use tone on almost every syllable which affects lexical meaning. For instance, the syllable 'ma' is polysemous, depending on the tonal contour it has four different meanings (mother, numb, horse and scold) in the Mandarin dialect (Kuo et al., 2015). Intonation languages are subdivided into two categories: pitch accent languages for which the placement of the tonal contrast is obligatory, such languages are Japanese, Turkish and Basque, and lexical stress languages for which pitch movement is not necessarily used to accompany stressed syllables, such languages are English, Spanish, German. The second function of intonation is the grammatical one that is used to disambiguate meaning at the syntactic level by for instance marking the boundaries of a sentence or highlighting the prominence of an element (focus). Intonation is also used in discourse to mark different kinds of sentences like questions or requests and to provide cues to regulate speech in turn-taking scenarios. Lastly, the paralinguistic function of intonation entails conveying information on the attitude and state of the speaker (Bolinger, 1982; Grice & Baumann, 2007; Nolan, 2020).

1.1.4. Intonation – Experimental evidence

Empirical work has showed that intonation is important in early language acquisition. Infants are naturally attracted to infant-directed speech (IDS) very early on in life (Cooper & Aslin, 1990; Fernald & Kuhl, 1987). This register is characterized by hyperarticulated vowels, slower speech rate than adult-directed pitch, but most importantly by exaggerated use of pitch. Phoneme acquisition and social preferences are positively affected by infant-directed speech (Adriaans & Swingley, 2017; Golinkoff et al., 2015; Schachner & Hannon, 2011 for a review see Cristia, 2013). Infants also benefit from intonational prominence to bootstrap syntax. French newborns distinguished sentences in French and Turkish based only on prosodic prominence. Both languages are syllable-timed and are very similar phonologically, in terms of word stress, syllable structure and vowel reduction, but French is a head-initial language, whereas Turkish is head-final. This difference is captured by prominence inside the utterance, which was enough of a cue for 2–3-month-olds to discriminate the sentences (Christophe et al., 2003). A similar tuning phenomenon, as perceptual narrowing for phonemes, has been observed for intonational markings of phrase boundaries. Infants at 4 months of age were sensitive to all three types of information that can be used to mark boundary clauses, i.e. pause, pitch and vowel duration. However, by 6 months, American English infants predominantly used the native cue, which is pitch (Seidl, 2007; Seidl & Cristià, 2008).

The contribution of intonation to language discrimination has been somewhat overlooked since the main focus of such research has been rhythm. Ramus (2002) was the first to disentangle the role of segmental, rhythmic and intonational cues in language discrimination with French newborns. Participants were presented with resynthesized sentences in Dutch (stress-timed language) and Japanese (mora-timed language) with different types of information available (see table 1 above). When participants were tested with no informative intonational cues (French contours were superimposed on both languages) in the 'sasasa artificial intonation' condition, neonates could no longer discriminate between the two languages. However, when participants were tested with 'saltanaj' with the same French contours superimposed, discrimination was observed only in the 1 minute after the switch of language but not when the analyses were conducted on the overall 2 minutes as in the rest of the experiments. The author discussed different interpretations of the results. He reported that the 'sasasa' stimuli might not be appropriate for infants. Ramus also discussed that rhythm might be less salient when intonation is removed, which could have led to the weaker discrimination (observed in the saltanaj with artificial intonation). He concluded that no firm conclusions can be drawn from the reported experiments and that future research should replicate these findings.

Intonation has also been proposed to play a role in within class discrimination. As already described, Nazzi et al. (2000) found that 5-month-old American-English infants can discriminate between their native dialect and the British one. The authors entertained the hypothesis that participants might have used differences in their pitch contours. Although this was not experimentally tested, they used natural speech providing infants with both prosodic and segmental cues. They concluded based on their acoustic analyses that rhythm and potentially other prosodic cues could have been contributed to discrimination. The role of intonation on within class discrimination was tested directly more recently. American English 5- and 7-month-old infants were tested using sentences in English and German. The results showed that this within class discrimination was achieved only at 7 months and that intonation was necessary for the participants to succeed in it (Chong et al., 2018). The lack of significant discrimination at 5 months of age is surprising as infants of the same age can discriminate Dutch from

British English or American English from British English (Nazzi et al., 2000). The authors tackled the issue by acoustically analyzing their stimuli. They found that their American English and German stimuli differed in subtler cues than the stimuli used in Nazzi and colleagues' work.

The use of intonation in adult language discrimination has received more attention. This research has primarily used two different types of tasks: language discrimination and language identification (LID). Other studies that can be perceived as language discrimination are neuroimaging investigations on neural entrainment to native and non-native languages[4]. Previous research has used different techniques to reduce segmental information and leave the prosodic one intact. Examples of such techniques are low-pass filtering, laryngograph signal, electronic monotonization and pulse trains. However, previous literature did not always disentangle the contribution of rhythm and intonation as they used languages that differ in both dimensions. Such research has investigated for instance the identification of American English, Japanese or Cantonese Chinese (Ohala & Gilbert, 1979; for similar results on French and English see (Maidment, 1983). In contrast, Komatsu et al. (2004) edited sentences in Japanese (tone lexical accent), English (stress lexical accent), Cantonese Chinese (accent lexical accent) and Spanish (stress lexical accent) in a way that only intonation was simulated. Critically, the first three languages use different types of lexical accent, whereas Spanish and English use the same. Participants could only identify sentences above chance when they belonged to languages that use different lexical accents (for a review on LID see Komatsu, 2007). Similarly, Ramus & Mehler (1999) tested the contribution of intonation alone on the ability of French adults to identify Japanese and English ('aaaa'). They found that intonational differences were not enough for adults to distinguish the two languages. In an attempt to test if nativeness with one of the languages tested affected the discrimination pattern, an additional group of American English participants was tested in the same condition. The results showed that American participants could identify the two languages by using only intonation.

[4] A proper description of such research falls outside the scope of the present investigation (e.g. Etard & Reichenbach, 2019; Lidji et al., 2011; Peña & Melloni, 2012)

However, participants were told in advance that they would listen to sentences in English and in a foreign language. This is not entirely equivalent to the task that French participants were tested in, since they were told they would listen to sentences from two exotic languages, confounding the implication of nativeness and that of explicit knowledge of the language tested (Ramus & Mehler, 1999). Further evidence testing French adults seem to suggest that intonation is not necessary for between class discrimination as they succeeded in discriminating English from Spanish and Catalan from English using only rhythmic cues. However, rhythmic cues were not enough for participants to perform finer, within class distinctions (English-Dutch, Catalan-Spanish)(Ramus et al., 2003). Thus, it is hard to draw any firm conclusions on adults' ability to discriminate between languages that differ prosodically using intonational cues.

The role of intonation on within class discrimination has also been investigated. Such research was conducted on adults discrimination capacities between Arabic and English (Moftah & Roach, 1988), different dialects of English (Bush, 1967; Richardson, 1973) ,dialects of Arabic (Barkat et al., 1999) and French dialects (Menard et al., 1999). Participants had access only to prosodic information which can be taken as evidence that intonation is an important cue, since rhythmic cues are not informative. A more recent study tested the contribution of intonation and rhythm combined (low-pass filtered speech), and separately by resynthesizing the sentences. Participants discriminated American English from German in all three conditions, whereas they discriminated American and Australian English only in the low-pass filtered and only rhythm condition above chance. Although, it is noteworthy that discrimination scores were relatively low in all three conditions when compared to previous literature (Vicenik & Sundara, 2013). These results corroborate previous findings using speech resynthesis on the necessity of intonation to discriminate within the same class using extrapolation since participants were unable to discriminate within the same class when they had no access to intonation cues (Ramus et al., 2003). The lack of discrimination in the only intonation condition for the two English dialects (Vicenik & Sundara, 2013) could be due to their acoustic similarity or the nature of the 'aaaa' resynthesis.

1.1.5. Conclusions – Language discrimination

The contribution of prosodic information to language discrimination has been tested extensively. Previous research has shown that rhythm is crucial to enable infants, adults and non-human animals to discriminate between languages belonging to different rhythmic classes. However, the cues that infants use to succeed in within class rhythmic discrimination remain underspecified. Previous research has focused on rhythm and has evaded answering this question with empirical evidence. Nazzi et al. (2000) claimed that infants most likely use prosodic information for within class discrimination, which is rather vague since prosody includes both rhythm and intonation.

The contribution of intonation to language discrimination is not clear. Data with infants is very limited. Previous research has shown that intonation is potentially not necessary for between-class discrimination in newborns, while it is essential for within class discrimination at 7 months. Such conclusions are based on just two published articles. Language discrimination based on intonational cues has been investigated more with adults. Data on between-class discrimination are hard to interpret since the role of rhythm and of intonation has not been always separated. The findings from Ramus et al., (2003) and Ramus & Mehler (1999) indicate that intonation is not necessary for adults to discriminate between languages of different rhythmic classes. Although, it might be enough for native participants to discriminate their native language from a rhythmically dissimilar one. In contrast, a clearer pattern is observed for within class discrimination. Adults use intonational cues to discriminate between languages that are rhythmically similar. Though, in some cases the sample size is extremely small (n=12) which suggests that further research needs to be done.

A summary of the results of language discrimination reported in the first section of the thesis can be found in table 2 for infants, table 3 for adults and table 4 for language identification with adults. Observing table 2, we can see that newborns used rhythmic cues to discriminate between languages and failed to discriminate within rhythmic classes. At 2 to 3 months of age results were mixed. Starting from 4.5 months of age, infants discriminated between language pairs that belong to the same rhythmic class, suggesting that they might have used other cues such as segmental or intonational one. Table 3 and Table 4 show that adults used rhythmic information to discriminate between rhythmic classes, while they used intonational ones for within class discriminations.

Lastly, previous research has not considered the potential contribution of phonetic information in early infancy as opposed to adulthood. The reason behind it being that there is no evidence that infants before 6 months of age have any phonemic knowledge. The next section will focus on the development of phonetic categories.

Table 2 [5]. Summary of previous results on Language Discrimination with infants							
Experiment	Language pair	Manipulation	Age of participants	Segmental diffs	Rhythmic diffs	Participants' L1	Discrimination
Minai et al. (2017)	English, Japanese	Natural	32-39 weeks (fetuses)	No/Limited	Yes	English	Yes
Mehler et al. (1988)	French, Russian	Natural	Newborns	Yes	Yes	French	Yes
		Low-pass filtered	Newborns	No/Limited	Yes	French	Yes
Moon, Cooper, & Fifer, (1993)	English, Spanish	Natural	Newborns	Yes	Yes	English/Spanish	Yes
Nazzi, Bertoncini, & Mehler, (1998)	English, Japanese	Low-pass filtered	Newborns	No/Limited	Yes	French	Yes
	English, Dutch			No/Limited	No		No

[5] In table 2 we summarize previous findings on language discrimination with infants. Each academic article is mentioned in the first column organized by ascending age of participants. The languages tested are presented in the second column, the nature of the stimuli in the third, the age of the participants in the fourth column. Whether the pair of language(s) tested differed at the segmental or rhythmic level (taking into consideration the nature of the stimuli) is summarized in column five and six respectively. The native language of the participants is written in column seven and whether discrimination was observed in the column eight.

	English +Dutch, Italian + Spanish			No/Limited	Yes		Yes
	English +Italian, Spanish + Dutch			No/Limited	No		No
Ramus, (2002)	Dutch, Japanese	Natural	Newborns	Yes	Yes	French	No
		Saltanaj		No	Yes		Yes
		Sasasa artificial intonation		No	Yes		No
		Saltanaj artificial intonation		No	Yes		Yes (only in the 1st minute post-switch)
Christophe & Morton, (1998)	French, Japanese	Natural	2 mo	Yes	Yes	English	No
	English, Dutch	Natural		Yes	No		Yes (marginally)
	Dutch, Japanese	Natural		Yes	Yes		Yes (marginally)

Cristia et al. (2014)	Quebecois, Parisian French	Audiovisual natural	3 mo	Yes	No	French	No
			5 mo	Yes	No		Yes
Dehaene-Lambertz & Houston, (1998)	English, French	Natural	2 mo	Yes	Yes	English	Yes
		Low-pass filtered		No/Limited	Yes	English	Yes
		Natural		Yes	Yes	French	No
		Low-Pass filtered		No/Limited	Yes	French	No
Kitamura, Panneton, & Best, (2013)	Australian, American English	Natural	3 mo	Yes	No	Australian English	Yes
Molnar, Gervain, & Carreiras (2014)	Basque, Spanish	Low-pass filtered	3.5 mo	No/Limited	No	Basque	Yes
						Spanish	Yes (only when habituated to Spanish)
						Basque & Spanish	Yes
Christophe et al.	French, Turkish	Synthetic sentences	1.5- 3 mo	No	No	French	Yes

(2003)							
Bosch & Sebastian-Galles, (1997) (preference)	Catalan/Spanish, English	Natural	4.5 mo	Yes	Yes	Catalan/ Spanish	Yes
	Catalan, Spanish	Natural		Yes	No		Yes
	Catalan, Spanish	Low-pass filtered		No/Limited	No		Yes
	Catalan/Spanish, English	Natural	4.5 mo	Yes	Yes	Catalan & Spanish	Yes
	Catalan, Spanish			Yes	Yes		No
	Catalan/Spanish, Italian			Yes	Yes		Yes
Bosch & Sebastian-Galles, (2001)	Catalan, Spanish	Natural	4.5 mo	Yes	No	Catalan/Spanish	Yes
	Catalan, Spanish	Natural		Yes	No	Catalan & Spanish	Yes

Butler et al. (2011)	South-West (native), Welsh English accent (unfamiliar)	Natural	5 mo	Yes	No	English	Yes
	Welsh, Scottish (unfamiliar English accents)	Natural	5 mo	Yes	No		No
Nazzi, Jusczyk, & Johnson, (2000)	British English, Japanese	Natural	5 mo	Yes	Yes	American English	Yes
	Italian, Japanese			Yes	Yes		Yes
	Italian, Spanish			Yes	No		No
	British English, Dutch			Yes	No		Yes
	British English, American English			Yes	No		Yes
	German, Dutch			Yes	No		No
Peña, Pittaluga, &	Spanish, Italian[6]	Natural	3 mo	Yes	No	Spanish	No

[6] Participants were presented with all three languages, but we have organized them in pairs for visualization purposes.

Mehler, (2010)			6 mo				Yes
	Spanish, Japanese		3 mo	Yes	Yes		Yes
			6 mo				Yes
	Italian, Japanese		3 mo	Yes	Yes		Yes
			6 mo				Yes
Chong, Vicenik, & Sundara, (2018)	American English, German	Natural	5 mo	Yes	No	American English	No
			7 mo				Yes
		Low-pass filtered	7 mo	No/Limited	No		Yes
		Natural with no intonation (monotone)	7 mo	Yes	No		No

Table 3. [7] Summary of previous results on Language discrimination with adults						
Experiment	Language pair	Manipulation	Segmental diffs	Rhythmic diffs	Participants' L1	Discrimination
Ramus, Dupoux, & Mehler (2003)	English, Spanish	Resynthesis only rhythm (flat sasasa)	No	Yes	French	Yes
	Catalan, English		No	Yes		Yes
	English, Dutch		No	No		No
	Catalan, Spanish		No	No		No
Vicenik, & Sundara (2013)	American English, German	Low-pass filtering	No/Limited	No	American English	Yes
		Resynthesis-Only rhythm	No	No		Yes
		Resynthesis-Only intonation	No	No		Yes

[7] In table 3 we summarize previous findings on language discrimination with adults. Each academic article is mentioned in the first column organized by alphabetical order. The languages tested are presented in the second column, the nature of the stimuli in the third. Whether the pair of language(s) tested differed at the segmental or rhythmic level (taking into consideration the nature of the stimuli) is summarized in column four and five respectively. The native language of the participants is written in column six and whether discrimination was observed in the column seven.

	American English, Australian English	Low-pass filtering	No/Limited	No	American English	Yes
		Resynthesis-Only rhythm	No	No		Yes
		Resynthesis-Only intonation	No	No		No

Table 4. [8] Summary of previous results on Language identification with adults

Experiment	Languages	Manipulation	Segment diffs	Rhythm diffs	Participants' L1	Identification
Atkinson (1968)	English, Spanish	Low-pass filtering	No/Limited	Yes	-	Yes
Barkat, Ohala, & Pellegrino (1999)	Eastern Arabic, Western	Natural	Yes	No	Western Arabic	Yes
					Non-native speakers	Yes

[8] In table 4 we summarize previous findings on language identification with adults. Each academic article is mentioned in the first column organized by alphabetical order. The languages tested are presented in the second column, the nature of the stimuli in the third. Whether the pair of language(s) tested differed at the segmental or rhythmic level (taking into consideration the nature of the stimuli) is summarized in column four and five respectively. The native language of the participants is written in column six and whether identification was observed in the column seven.

<table>
<tr><td rowspan="2"></td><td rowspan="2">Arabic</td><td rowspan="2">Sinusoidal pulses</td><td rowspan="2">No</td><td rowspan="2">No</td><td>Western Arabic</td><td>Yes</td></tr>
<tr><td>Non-native speakers</td><td>No</td></tr>
<tr><td>Bush (1967)</td><td>American, British, Indian English</td><td>Low-pass filtering</td><td>No/Limited</td><td>No</td><td>American English</td><td>Yes</td></tr>
<tr><td rowspan="6">Komatsu, Arai, & Sugawara (2004)</td><td>Chinese, English</td><td rowspan="6">Pulse train simulating f0</td><td rowspan="6">No</td><td rowspan="6">No</td><td rowspan="6">-</td><td>Yes</td></tr>
<tr><td>Chinese, Japanese</td><td>Yes</td></tr>
<tr><td>Chinese, Spanish</td><td>Yes</td></tr>
<tr><td>English, Japanese</td><td>Yes</td></tr>
<tr><td>English, Spanish</td><td>No</td></tr>
<tr><td>Japanese, Spanish</td><td>Yes</td></tr>
<tr><td>Maidment (1983)</td><td>French, English</td><td>Laryngograph waveform</td><td>No</td><td>Yes</td><td>English</td><td>Yes</td></tr>
<tr><td>Menard, Ouellon, & Dolbec (1999)</td><td>European French,</td><td>Low-pass filtering</td><td>No/Limited</td><td>No</td><td>Quebecois French</td><td>Yes</td></tr>
</table>

	Quebecois French					
Moftah, & Roach (1988)	Arabic, English	Laryngograph waveform	No	No	Arabic or English	Yes
		Low-pass filtering	No/Limited	No	Arabic or English	Yes
Muthusamy, Barnard & Cole (1994)	English, Farsi, French, German, Japanese, Korean, Mandarin Chinese, Spanish, Tamil, Vietnamese	Natural	Yes	Yes	10 English speakers and 2 from each language other than English	Yes
Ohala & Gilbert (1979)	English, Cantonese Chinese, Japanese	Triangular pulse train	No	Yes	English, Chinese, Japanese	Yes
Ramus & Mehler (1999)	English, Japanese	Resynthesis (‘aaaa’)	No	No	French	No
					American English	Yes

1.2. Establishment of phonemes

In the first months of life, infants can discriminate between virtually all sounds regardless of whether they are found in their language of exposure or not. This ability has been observed by testing 1-4 month-old infants using consonants as well as vowels (Cheour-Luhtanen et al., 1995; Eimas et al., 1971; Trehub, 1973, 1976; for a review see Aslin et al., 1998; Werker & Gervain, 2013). In the following months, infants' ability to discriminate non-native contrasts starts to decline, while simultaneously they improve in distinguishing phoneme contrasts present in their language of exposure. This phenomenon has been defined as *perceptual narrowing*. The most common way to study this attunement is by testing infants' abilities to discriminate pairs of sounds that might be part of the phonetic repertoire of their language of exposure or not. Perceptual narrowing is evident when infants no longer discriminate a phonetic contrast that does not exist in their language environment.

Werker & Tees (1984) were the first to provide evidence of perceptual narrowing. Infants were tested at three different age groups, 6-8, 8-10 and 10-12 months to determine when the decline starts. Two pairs of non-native speech contrasts were used. The first one was /k'i/-/q'i/ from Thompson, i.e. an Interior Salish language (Native Indian), and the second one was /ʈa/-/t̪a/ from Hindi. English infants belonging to the two youngest groups, 6-8[9] and 8-10, distinguished between both contrasts but failed to do so at 10-12 months, showing that the decline in discriminating non-native consonantal contrasts emerges at the end of the first year of life. Diachronically, this decline is evident with adult participants, too. English adults were unable to discriminate the Thompson contrast, while native speakers had no trouble in doing so.

Language experience has been found to modulate perception of speech segments differently. Perceptual narrowing has been observed earlier for vowels, already at the 6th month of life, which is

[9] The data on the Hindi contrast with 6-8 month old English infants were collected in a previous experiment (Werker et al., 1981).

a few months earlier than the first signs of perceptual narrowing for consonants occur (Kuhl et al., 1992; Polka & Werker, 1994; for a review see Tsuji & Cristia, 2013). The first signs of perceptual narrowing for lexical stress and tone take place even earlier, at 5 months (Yeung et al., 2013). The attunement to the native repertoire has been replicated widely using a variety of techniques, both behavioral and neuroimaging, EEG and MEG measurements (Bosch & Sebastián-Gallés, 2003; Bosseler et al., 2013; Cheour et al., 1998; Ortiz-Mantilla et al., 2013). This perceptual reorganization is not only affected by language exposure but also by maturational brain constraints. Peña et al. (2012) showed that infants' discrimination capacity declined for non-native contrasts, lower amplitude in the mismatch response, at the same maturational age between preterm and full term, i.e. at 12 months of age, indicating that having additional experience to broadcast speech does not accelerate perceptual narrowing. Once infants have attuned to the phonetic repertoire and prosodic patterns of their native language, other knowledge can be scaffolded such as phonotactics and grammar.

1.2.1 Vowel perception

Infants acquire information on the native phonetic repertoire during the first year of life. The phonetic repertoire consists of both vocalic and consonantal categories, which comprise of different realizations of speech sounds. The classification of a speech sound into a different category depends on its functional relevance, i.e. conveys a different meaning. The acquisition of the native phonetic categories is a true feat, due to the high variability of the speech signal. Infants need to create them in face of great variability. Sounds change depending on the speakers' identity in terms of age or gender and depending on the position they are found in a word. Research has shown that infants are able to discriminate between variants of different realizations of /a/ and /i/ from 6 months of age, indicating language constancy (Kuhl, 1979).

Across languages vowels tend to be less in number than consonants, which can be described in terms of the consonant-vowel ratio. (Maddieson, 2013). For instance, in Spanish there are 3.8 more consonants than vowels. Though, vowels are very frequent in the speech signal, for example in Spanish they occupy 43.8% of the speech signal (Ramus et al., 1999) Vowels carry most of the prosodic information, which as we have mentioned in the previous section is an integral part of infants’ speech perception during the first months of life. The acoustic saliency of vowels, longer in duration and more stable, results in them being perceived better in utero (Granier-Deferre et al., 2011; Querleu et al., 1988). Vowels are also used during the first months of life to memorize new words and encode relationships between an audiovisual referent showing an action or an object (Jara et al., 2021). Five-month-old infants are sensitive to mispronunciations of their name only if it is related to a vowel (eg. Elix instead of Alix) (Bouchon et al., 2015). Therefore, vowels attract infants’ attention first.

Infants perceive vowels differently depending on their quality, i.e. being an ideal exemplar of its speech category or not when presented with similar sounds (Kuhl, 1991). For instance, when infants were presented with an ideal referent of the sound /i/ against its surrounding realizations, they found them as being more similar to when they were presented with a non-ideal referent of the same category. These results suggest that the ideal referent of each speech category acts like a magnet to its surrounding sounds. Interestingly, at 6-months, *the perceptual magnet effect* is evident only for sounds found in the native language (Kuhl et al., 1992). American infants perceived the native /i/ as identical to its variants significantly more than when they were presented with the nonnative Swedish /y/. The reverse pattern was observed with Swedish infants, who showed the perceptual magnet effect only for the native /y/. These results indicate the infants have established some phonetic knowledge. These attracting effects of prototypes on allophones have been described as the perceptual magnet effect which gave birth to the Native Language Magnet theory (NLM) (Kuhl, 1993). There is some evidence consistent with the magnet effect in newborns but these results have never been replicated (Moon et al., 2013).

A meta-analysis on vowel perception during the first year of life corroborates that starting at 6 months of life, infants exhibit the first signs of perceptual narrowing for vowels (Tsuji & Cristia, 2013). Native language perception of speech contrasts starts at 6 months, depending on the task and contrast. Cheour et al., (1998) provided corroborating electrophysiological evidence of the emergence of vowel categories in Finnish-learning infants. These investigators tested 6- and 12-month-old infants with an odd-ball paradigm. The standard (/e/) and one of the two deviants (/ö/) were native phonemes. The other deviant was an Estonian phoneme (/õ/). Infants at 6 months showed a Mismatch Negativity effect for both deviant sounds, whereas at 12 months the MMN was significantly bigger for the native vowel. The discrepancy on the age of the language specific effects on vowel perception between Cheour et al. and Kuhl et al. (1992) could be due to several reasons. The vowel inventories of the three languages, i.e. Finnish, American English and Swedish are large (Maddieson, 2013; van et al., 1994), so the number of vowels cannot account for the differences. However, vowel systems differ not only in terms of quality but also in terms of vowel length. Vowel length is phonemic in Finnish but not in English. This difference is related to the hypothesis that vowel length is acquired first by infants than vowel quality (Paillereau et al., 2021). It could also be that the oddball paradigm is not as sensitive as the head-turn preference procedure to detect an early change in behavior due to the native language. This paradigm has also proven insensitive to detect asymmetries in vowel perception of adults, an effect that has been found behaviorally (Polka et al., 2021). The MMN effect is thought to be an early pre-attentive response. Polka et al. (2021) found that time-frequency analysis is more appropriate. Lastly, it could be that the explored vowels in each language are not of similar frequency.

Infants’ ability to discriminate between native vowels is not only affected by the presence or absence of sounds in their native language but also by their position in the acoustic space. This space is defined in most languages by two formants that carry the acoustic energy. The first formant (F1) is inversely related to the height of the tongue during articulation and the second formant (F2) related to the backness of the vowel. This organized vowel space results in some vowels being peripheral and others more central. Previous research has shown that there are directional asymmetries for both infants and adults, being easier to discriminate a central from a peripheral sound (Polka et al., 2021; Polka & Bohn, 1996, for a review see Polka & Bohn, 2003). Peripheral sounds act as perceptual attractors rendering discrimination harder. The acoustic salience of peripheral sounds becomes susceptible to the native language as infants grow. Catalan and Spanish infants at 4 and 6 months were only able to discriminate /e-i/ (I is the peripheral vowel) and not in the opposite direction (Pons et al., 2012). When they were habituated with the peripheral sound as a referent, then in the test phase they could not tell apart the two sounds. At 12 months instead of observing this acoustic directionality, a different pattern of results was found for each population based on the frequency of the sounds. In Catalan /i/ is more frequent than /e/ (Rafel, 1980), whereas in Spanish the opposite holds true (Alcina & Blecua, 1975). This difference in frequency rendered the /e-i/ discriminable only by Catalan infants and the /i-e/ discriminable only by the Spanish infants. Frequency of appearance has also been proposed as a factor in the order that non-native contrasts are lost. English learning infants were tested at 6.5 and 8.5 months of age on whether they can discriminate two non-native contrasts (dorsal (/k’ɜ/-/q’ɜ/), coronal (/ʈɜ/-/ [t̪]) consonant pairs). These categories are of different incidence in English with coronal stops being more frequent. The younger group of participants discriminated the two contrasts equally well, while the older group was worse in the coronal contrast, indicating a parallel process in which frequent phoneme categories are acquired first and consequently, non-native contrasts of the same category are lost earlier (Anderson et al., 2003).

The research just reviewed about perceptual learning provides important information on the time-course of infants’ tuning to the native phonetic repertoire, but they are silent with regards to the learning mechanism. One such mechanism that has been proposed to be behind phoneme acquisition is distributional learning.

1.2.2 Distributional learning – Phoneme categories

Distributional learning has been proposed as the functional mechanism underlying phoneme acquisition. Maye et al., (2002) tested whether the distribution of sounds, i.e. their frequency of appearance, would modulate the ability of 6 and 8 month-old infants to discriminate a phonetic contrast. They created 8 different tokens forming a continuum of [da] to [ta] stimuli, the former being a voiced unaspirated stop consonant and the latter a voiceless unaspirated stop consonant. Participants were presented with either a bimodal or unimodal distribution of the different tokens, depending on the frequency of each token. Infants that heard more exemplars from the end of the continuum were exposed to a bimodal distribution, whereas participants that heard exemplars from the middle were exposed to a unimodal distribution (figure 1.3.). Therefore, bimodal distribution induced the formation of two phonetic categories, and the unimodal the formation of one. During the test infants were presented with alternating, two different stimuli, and non-alternating, same stimulus, trials. If infants had detected the difference across the trials, they were expected to have statistically different looking times. The findings showed that participants discriminated the test stimuli only after having been exposed to the bimodal distribution. Similar results were obtained by testing adults using a voice onset time (VOT) continuum /d~D/. Adults were able to discriminate only after having been exposed to the bimodal distribution. However, this distribution effect was not generalizable to the same kind of contrast, but with a distinct place of articulation /g-G/. The results indicated that participants failed to discriminate above chance for the untrained contrast (Maye & Gerken, 2001), while there is evidence that infants can generalize in a similar task (Maye et al., 2008).

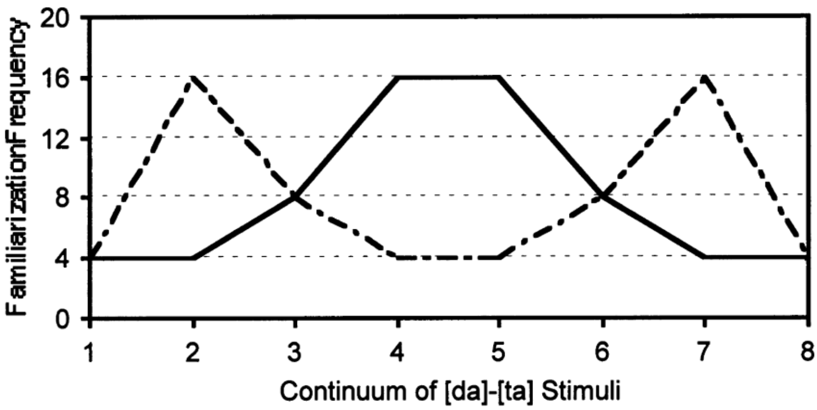


Figure 1.3. Figure reproduced from Maye et al. (2002) showing the Unimodal (solid line) and Bimodal (dotted line) distributions that infants heard during familiarization. The y axis stands for the frequency each token was heard during the experiment.

The universality of a statistical learning mechanism led Maye and colleagues (2002) to suggest that it could be found in younger infants than 6-month-olds. They even suggested that the time in between the first evidence of this mechanism (6 months) and the first signs of perceptual narrowing for consonants (10 months)

(Werker & Tees, 1984) could be due to the need to extract frequency of occurrence from natural input, which is not as concentrated as in a laboratory setting. Therefore, distributional effects on the discrimination of contrasts should be evident earlier for vowels since they are less in number. This hypothesis was put to the test when 2–3-month-old Dutch infants were tested on their perception of a vowel contrast (/æ/~/ɛ/). Due to the young age of the participants, they were tested using the oddball paradigm. Infants that were familiarized to a bimodal distribution were better at discriminating the vowel contrast, i.e. larger mismatch response (Wanrooij et al., 2014). Limitations to distributional learning might exist from a certain age onwards. English 10-11-month-old infants were unable to discriminate a non-native consonant contrast in voicing and place of articulation, regardless of distributional information. This failure to discriminate was overturned when the duration of familiarization was doubled (Yoshida et al., 2010)[10]. Corroborating evidence on the declining efficacy of this mechanism through the first year of life comes from research conducted on a native contrast (/ra/~/la/) at three different ages, 5, 9 and 12 months of age. They found that the influence of distributional information started declining from 9 to 12 months of age (Reh et al., 2021; for similar results on tone discrimination Liu & Kager, 2017;). Various attempts have also been made to simulate infants' phonetic learning using computational models (Adriaans & Swingley, 2012, 2017; McMurray et al., 2009; Schatz et al., 2021; Vallabha et al., 2007)

The literature on adults' use of distributional learning is less conclusive. There is some evidence that distributional information helps adults to discriminate non-native vowel pairs. Spanish participants discrimination of /a/~ /ɑː/ improved after having been trained in the enhanced bimodal distribution, i.e. the peaks of the distribution were closer to the endpoints of the continuum than the typical bimodal distribution (Escudero et al., 2011). On a similar task, Bulgarian participants' perception improved of two non-native vowel contrasts (/a/~/ɑ/, /i/~/ɪ/) after unsupervised distributional training (Gulian et al., 2007). In contrast, no such facilitatory effects were observed when Dutch participants were tested behaviourally (Wanrooij, 2015) or in a neuroimaging task (Wanrooij et al., 2014b). This different pattern of results could be related to the

[10] This was only tested with the place of articulation contrast (/da/~/ta/).

structure of the vowel space of each language. Wanrooij (2015) suggested that the number of vowels might be the reason why distributional learning was observed only for Spanish and Bulgarian participants. They have a few vowels, 5 and 6 respectively, whereas Dutch has 15. Therefore, Dutch participants probably had to change a native vowel boundary in order to discriminate the English pair. In the case of Spanish and Bulgarian, participants had to create a boundary. Therefore, the mechanisms underlying distributional learning for adults are different than the ones in infancy. Adults have a fully developed phonetic space in place for at least one language. By exposing them to lab-induced distributions of sounds, they are implicitly asked to 'reorganize' their existing phoneme categories. This reorganization is difficult due to the loss of plasticity that very young infants have before perceptual narrowing.

1.2.3 Conclusions – Establishment of phonemes

Summing up, during the first year of life infants reorganize their speech categories. Infants transform from universal listeners to listeners of their native language. Perceptual narrowing is a developmental process in which infants first tune to the native vocalic system at 6 months (Kuhl et al., 1992; Polka & Werker, 1994) and then do the same for consonants at 10 months of age (Werker & Tees, 1984). The proposed mechanism behind the tuning process has been described as distributional learning (Maye et al., 2002, 2008). However, previous research has investigated the establishment of phonemes mainly in the second semester of life. This is due to the results of discrimination tasks. Infants' equivalent capacity to discriminate non-native and native vowel contrasts has been interpreted as evidence that infants possess no phonetic knowledge yet. Although this result has been replicated widely it does not mean that infants have no phonetic knowledge.

It seems unlikely that phoneme categories emerge in a relatively short period of time (or "abruptly"). Maye et al. (2002) suggested that the time elapsed between the effects of distributional learning at 6 months and the first signs of perceptual narrowing for consonants at 10-12 months is probably due to the need to accumulate enough exposure to the native language in order to build distributional consonantal representations. Following that argument, they said that supportive evidence of this process comes from vowel perception. Infants attune to the native vowel repertoire first because they are less in number, and they occupy most of the speech signal. Therefore, infants can create phonemic categories faster based on the input from their environment. Converging evidence comes from research showing that 2-3-month-old infants' perception of a vowel pair is modulated by distributional information (Wanrooij et al., 2014). These authors suggest that infants before the first signs of perceptual narrowing have started building phonemic representations. This is the central goal of the present dissertation: to investigate the precursors of the establishment of the phoneme repertoire.

1.3. Outline of the current thesis

The emergence of phonetic categories has been observed in the second semester in life. One potential mechanism underlying it is distributional learning. Given that distributional learning influences discrimination of vowels as early as 2 to 3 months of age and the first signs of perceptual narrowing for vowels occur at 6 months, it is possible that infants develop some kind of "proto-segmental" information of their native vowel space. This information could be on the distribution of (uncategorized) sounds naturally found in their native language. We tested this hypothesis in two series of experiments (chapters two and four), taking advantage of relevant differences in the distribution of vowels in three different languages[11], Eastern Catalan, Western Catalan and Spanish. All three are Romance and syllable-timed languages. Importantly, they are different regarding their vowel incidence. Eastern Catalan has vowel reduction. i.e. all mid vowels are reduced when they are not in a stressed position, which results in low vowels being very frequent (61%), while mid vowels are not as frequent (16%) (Rafel, 1980). Western Catalan and Spanish do not have vowel reduction which results in them having the opposite frequency patterns. Mid vowels in Spanish are more frequent (49%) than low vowels (29%) and the same can be observed for Western Catalan[12] (mid vowels: 46%, low vowels: 36%, see figure 1.4.) (Alcina & Blecua, 1975).

[11] We consider dialects as languages for the sake of simplicity.
[12] The vowel distribution of Western Catalan was calculated from the stimuli we used in Chapter 2.

Our aim was to investigate the presence of native language knowledge at the segmental level just before perceptual narrowing has been reported, we tested 4–5-month-old infants.

In chapter two, we investigated whether infants have some protosegmental information in place that would allow them to discriminate between languages. To test this, we combined the three languages into pairs, one that has different vowel distributions, Eastern and Western Catalan, and another pair with comparable vowel distributions, Western Catalan and Spanish. We report a series of 4 experiments testing infants' ability to discriminate Eastern and Western Catalan (with very different vowel distributions), and Western Catalan and Spanish (with very similar vowel distributions). We used these languages using natural and low-pass filtered sentences. We expected that infants would be able to discriminate only the first pair of languages due to their differences at the segmental level. When infants were presented with reduced segmental information due to low-pass filtering, such discrimination should no longer be observed. We found that 4.5-month-old infants could discriminate Eastern and Western Catalan only when the sentences were natural speech, suggesting that infants relied on segmental information. To our surprise, we also observed discrimination between Western Catalan and Spanish with natural sentences and marginal discrimination with low-pass filtered sentences. We attributed such effects to the differences in their intonational cues, as Catalan and Spanish have been reported to differ in their frequency of intonational patterns to mark boundaries or phrases.

Adult language discrimination is guided mainly by segmental information. It has been shown that they can also use prosodic information, but its efficacy depends on the languages compared and the familiarity the participants have to the languages tested. Chapter three investigates whether adults can use intonational cues to discriminate Western Catalan from Spanish. An additional group of non-native speakers was also tested to verify assess if familiarity with the languages at test would facilitate discrimination. We found that native and non-native adults are successful in discriminating languages using intonational cues. Nativeness had a modulating factor, since native speakers were better at the task. In general, the discrimination scores were far from being perfect, providing evidence on how important segmental cues are for adults.

Chapter four presents additional evidence on the language-specific knowledge of the vowel distribution in 4.5-month-old-infants. The design of the language discrimination experiments could not rule out that infants developed some temporary sensitivity to vowel distribution during the familiarization phase of the study. In chapter four, we tested infants' preference for lists of nonwords that mimic the vowel distribution of either their native language (Catalan or Spanish) without any previous familiarization. The results show that 4.5-month-old infants prefer listening to lists that resemble the vowel distribution of their native language. The results provide additional support to the hypothesis that before 6 months of age infants already have in place some coarse representation on the distribution of the vowels present in their native language.

Chapter five studies whether adults are also sensitive to the vowel distribution of their native language using the same stimuli as in chapter four. On average half of the words in the Spanish lists were phonotactically illegal in Catalan due to vowel reduction. Adult participants can use either vowel distribution (global level) or phonotactic knowledge (word level). If adults focused on the global level, we expected to find a preference for the lists resembling the native vowel system, but if they focused on the word level, only Catalan participants would be able to distinguish the two types of lists. Our results support the second hypothesis. Only Catalan participants were able to detect the lists of words that resemble their native language potentially relying on phonotactic permissibility.

In chapter six, conclusions of this dissertation are presented, along with potential explanations of the effects found. We also discuss limitations of the experiments reported in this thesis and future directions that can complement our findings.

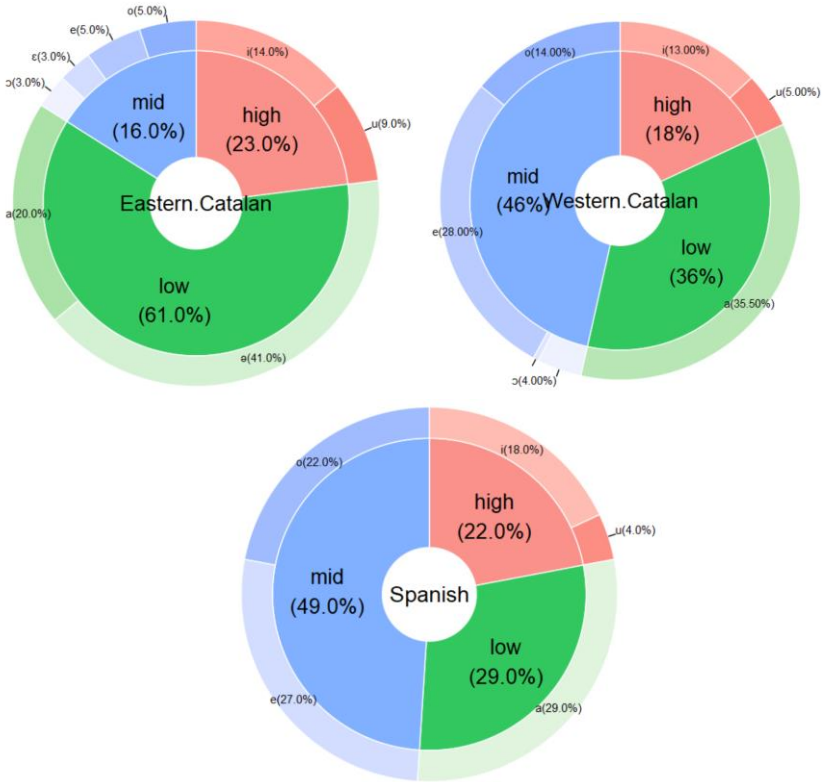


Figure 1.4. Vowel distributions of Eastern Catalan, Western Catalan, and Spanish. Individual vowels are represented in the outer layer of the plot and grouped percentages based on F1 are depicted in the inner layer.

Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. Cognition, February, 104628. https://doi.org/10.1016/j.cognition.2021.104628

2. The ontogeny of early language discrimination: Beyond rhythm

Cognition 213 (2021) 104628

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The ontogeny of early language discrimination: Beyond rhythm[☆]

Konstantina Zacharaki[*], Nuria Sebastian-Galles

Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, 08005 Barcelona, Spain

ARTICLE INFO

Keywords:
Language discrimination
Vowel distribution
Intonation
Infants
Bilingualism

ABSTRACT

Infants can discriminate languages that belong to different rhythmic classes at birth. The ability to perform within-class discrimination emerges around the fifth month of life. The cues that infants use to discriminate between prosodically close languages remain elusive. Segmental information could be a potential cue, since infants notice vowel mispronunciations of their names, show the first signs of word recognition and the first signs of perceptual narrowing for vowels around 6 months of age. If infants have in place some proto-segmental information, most likely it is about vowels. Another potential cue infants may use to discriminate languages is intonation. We tested participants using sentences in Eastern Catalan, Western Catalan and Spanish. The two Catalan dialects and Spanish belong to the same rhythmic class, they are syllable-timed, but they differ in terms of vowel distribution, given that only Eastern Catalan has vocalic reduction. The vowel distributions of Western Catalan and Spanish are more comparable. However, they differ in terms of their intonational patterns. In Experiment 1, we tested the ability of 4.5-month-old infants learning Eastern Catalan and/or Spanish to discriminate between sentences in Eastern and Western Catalan and in Experiment 2 their ability to discriminate between sentences in Western Catalan and Spanish. In order to disentangle the contribution of segmental and suprasegmental information, we also tested infants using low-pass filtered sentences in the two dialects (Experiment 3) and low-pass filtered sentences in Western Catalan and Spanish (Experiment 4). Infants discriminated the two Catalan dialects only when the stimuli were natural sentences, whereas they were able to discriminate between Western Catalan and Spanish when the stimuli were either natural or low-pass filtered sentences. The research also provides evidence of equivalent language discrimination abilities in infants growing up in monolingual and bilingual environments.

1. Introduction

During the first months of life, infants' linguistic knowledge quickly develops from some coarse language universal processing capacities at birth to the identification of some highly familiar words at 6 months. Research with young infants has shown that they are able to distinguish virtually any speech sound present in human languages, even if not produced or perceived by their parents (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) and they can discriminate between some human languages (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993; Nazzi, Bertoncini, & Mehler, 1998). Such capacities are universal as they do not appear to depend on previous exposure to any specific language and they are likely based on some ancillary perceptual mechanisms, which are common with other animals (Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, & Sebastián-Gallés, 2005). At 6 months, they are able to identify some highly frequent words (Bergelson and Swingley, 2012; Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Tincoff & Jusczyk, 1999). In spite of such remarkable development, our comprehension of how language develops during the first half of the first year of life is quite limited, at least in comparison to research investigating the second half of the first year of life. In the present research we test the hypothesis that before 6 months of age infants already have some coarse representation of the vocalic system of their native language and it is that exact knowledge which can be used to discriminate languages. We also provide evidence of the use of intonation in early language discrimination.

One of infants' initial processing capacities is neonates' ability to discriminate some languages from others, for instance, they can discriminate French from Russian, or English from Spanish (Mehler et al., 1988; Moon et al., 1993)). It has been proposed that early language discrimination is based on the computation of some prosodic

[☆] This paper is a part of special issue "Special Issue in Honour of Jacques Mehler, Cognition's founding editor".

[*] Corresponding author.
E-mail address: konstantina.zacharaki@upf.edu (K. Zacharaki).

https://doi.org/10.1016/j.cognition.2021.104628
Received 2 November 2020; Received in revised form 4 February 2021; Accepted 7 February 2021
Available online 19 February 2021

information, that is, on sensitivity to differences in the *rhythmic* and *intonational* aspects of languages.

Although there is no unique definition of linguistic rhythm (Allen & Hawkins, 1980; Fletcher, 2010; Nazzi & Ramus, 2003; Nespor, Shukla, & Mehler, 2011; Nolan & Jeon, 2014), it has been linked to alternations of periodically recurring strong and weak elements. Languages have been classified as stress-timed, syllable-timed or mora-timed languages depending on the recurring unit -foot, syllable or mora- at which periodicity is most prevalent (Abercrombie, 1982; Ladefoged, 1975; Lloyd James, 1940; Pike, 1945). Examples of languages of each group are English and Thai (stress-timed), Spanish and Korean (syllable-timed) and Gilbertese and Japanese (mora-timed). More recent conceptualizations consider that language rhythm is better captured in a continuum way, according to different phonological parameters, such as syllabic structure complexity, vowel reduction or the phonetic realization of stress (Dauer, 1983). Stress-timed languages tend to have greater variability and complexity of their syllabic inventory, and reduced vowels in unstressed positions. In contrast, syllable-timed languages allow simpler and fewer combinations of syllables, whereas vowels are usually realized in the same manner, regardless of stress. Several studies have shown that if two languages are sufficiently different in terms of their linguistic rhythm, newborns will be able to discriminate them (Byers-Heinlein, Burns, & Werker, 2010; Mehler et al., 1988; Moon et al., 1993; Nazzi et al., 1998; Ramus, 2002). They can even discriminate between pairs of language, i.e. a mix of English and Dutch sentences from a mix of Spanish and Italian, but not in any combination, they cannot discriminate mixing English and Spanish sentences from Italian and Dutch sentences (Nazzi et al., 1998). Newborns show equivalent patterns of language discrimination even when sentences have been manipulated and segmental information has been removed, either by low-pass filtering the sentences (Byers-Heinlein et al., 2010; Mehler et al., 1988; Nazzi et al., 1998) or by resynthesizing them and replacing the original phonemes by a set of language-neutral ones (Ramus, 2002). However, when prosodic information is removed, and stimuli have a flat intonation, discrimination disappears (Ramus, 2002). Therefore, the ability to compute some prosodic information seems to be an integral part of neonates' ability to discriminate between languages.

By 4–5 months of age, infants have gathered new information that will allow them to discriminate some rhythmically closer languages (*within-class language discrimination*). At this age, monolingual and bilingual infants, learning Catalan and Spanish, two Romance languages classified as syllable-timed (Prieto, Vanrell, Astruc, Payne, & Post, 2012; Ramus, Nespor, & Mehler, 1999) can discriminate them (Bosch & Sebastian-Galles, 1997, 2001). Analogously, 5-month-old infants learning American English can discriminate Dutch from British English, and their native dialect from British English (Nazzi, Jusczyk, & Johnson, 2000), but fail to discriminate within a nonnative rhythmic class (Italian vs. Spanish). They even fail to perform a native within-class discrimination (Dutch vs. German). Patterns of discrimination at this age are quite complex, showing partial within-class language discrimination, usually requiring familiarity with one of the languages of test for successful discrimination. Studies investigating dialect discrimination, that can be considered as a within-class language discrimination situation also point in the same direction. Butler, Floccia, Goslin, and Panneton (2011) observed that 5-month-old infants exposed to South-West-English could discriminate their dialect from an unfamiliar Welsh English accent.[1] However, they could not discriminate between Welsh and Scottish English, unfamiliar accents to them. One explanation for the increase in within-class discrimination abilities has been that by this age infants have developed the capacity to perform more fine-grained rhythmic computations focused on their native language and rhythmic class (Nazzi et al., 2000).

Although the evidence of the role of rhythm in early language discrimination capacities is abundant, it does not preclude that other types of prosodic information may have contributed to the observed patterns of results. The stimuli used in the studies just reviewed included most of the prosodic information, specifically intonational cues (including studies where sentences were manipulated either by low-pass filtering or by the resynthesis procedures described in Ramus and Mehler (1999). As languages differ in their melodic properties, intonation is likely to play a role in language discrimination as well (Maidment, 1976, 1983; Ohala & Gilbert, 1979). Supporting the relevance of different suprasegmental information, Nazzi et al. (2000) discussed intonational differences between American and British English that could explain how 5 month old infants could discriminated them. Also, White, Floccia, Goslin, and Butler (2014) concluded that language discrimination at 5 months relies on the use of localized gradient differences in prosody.

To our knowledge, the specific contribution of intonational information in early language discrimination has been targeted in just two different studies, one with newborns and one with 5- and 7-month-olds. In a series of experiments, Ramus (2002) tested French neonates' capacity to discriminate resynthesized Dutch and Japanese sentences in which different types of phonological information had been removed from them. Replicating previous results, he observed reliable discrimination when sentences preserved rhythmic and intonational information, but no discrimination was observed when sentences were resynthesized superimposing the intonational pattern of French sentences, therefore by neutralizing language-specific intonational information. Chong, Vicenik, and Sundara (2018) investigated the role of intonation in older infants. They compared the ability of American English learning infants to discriminate American English from German (two stressed-timed languages) when different types of information were removed from the signal. Contrary to Nazzi et al. (2000) who found within-class language discrimination at 5 months of age, Chong, Vicenik, and Sundara (2018) failed to observe discrimination between American English (their native language) and German in 5-month-olds in any of their experimental manipulations. The results of 7-month-olds showed discrimination whenever prosodic intonation was present: natural sentences and low-pass filtered ones. Crucially, 7-month-olds failed to discriminate when the original pitch contours were substituted by a monotone one; therefore, leaving rhythmic and segmental information, but removing the intonational one. The fact that Chong et al. (2018) did not find within-class language discrimination (between American English and German) at 5-months unlike Nazzi et al. (British English vs. Dutch), led them to compare the intonational patterns in both studies. Chong et al. (2018) measured several prosodic dimensions of Nazzi et al.'s Dutch and British English stimuli and concluded that stimuli in German and American English from Chong et al. (2018) were intonationally more similar than the ones used in Nazzi et al. (2000).

The pattern of results of early language discrimination studies has led researchers to favor prosodic information (rhythmic or intonational) over segmental information when explaining language discrimination in the first months of life. This is because patterns of language discrimination did not change when segmental information had been cancelled (Chong et al., 2018; Molnar, Gervain, & Carreiras, 2014; Nazzi et al., 2000; Nazzi & Ramus, 2003; Vicenik, 2011), but the removal of prosodic (rhythmic or intonational) information has generally resulted in diminished discrimination. Additionally, the investigation of language processing in the first 6 months of life suggests that infants do not possess a proper representation of the phonological properties of their language of exposure other than prosody, or at least that they do not use such information in laboratory tasks. There is no solid evidence that at this age infants use information about phonemes (Johnson, Jusczyk, & Ramus, 2003; Nazzi et al., 2000) or that they have any knowledge about the phonotactics of their native language (Friederici & Wessels, 1993;

[1] In the present text we will use the term "language" to refer to languages and dialects, as often the difference between typologically close languages and dialects is blurry.

Jusczyk, Cutler, & Redanz, 1993). In contrast, at 5 months of age infants can already recognize some familiar word forms (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Mandel, Jusczyk, & Pisoni, 1995; Mersad & Nazzi, 2012) and they can comprehend some words at 5–6 months (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). At this age infants show sensitivity to vowel (but not to consonant) mispronunciations of their own name (Bouchon et al., 2015). Taking into consideration these findings, it is likely that infants at 5-months have some proto-segmental information available which probably is related to vowels of their native language(s).

By 6 months of age infants have already established knowledge of the vocalic system of their language. There is evidence that at this age the perceptual vowel space has already "warped" to reflect the properties of their native language (Kuhl, 1991; see also Kuhl et al., 1992; Polka & Werker, 1994; Werker & Tees, 1984). The computation of distributional information is one of the powerful tools infants have at their disposal to establish the phoneme repertoire of their native language (Maye, Werker, & Gerken, 2002). If at 6 months of age, native vowels are already in place, it is likely that by 4–5 months of age infants possess some language-specific knowledge of the distribution of vocalic sounds in their own language. The hypothesis that young infants may be sensitive to vocalic information had already been put forward by Mehler and colleagues with their TIGRE proposal (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996). Though Ramus, Nespor, and Mehler (1999) considered rhythm as primary information for early language discrimination, they also proposed that 'the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalysed noise (i.e. consonants)' (Ramus et al., 1999, p. 270). The authors even suggested that infants may be sensitive to cross language differences in number and position of vowels. The central goal of the present research is to test this hypothesis. To this end we will investigate if 4–5-month-old infants can discriminate languages that differ primarily in the distribution of vowel sounds. We will take advantage of the differences in the distribution of vowels between Spanish and two dialects of Catalan: Eastern (Barcelona) and Western (Tortosa). We will also contribute to the rather sparse data investigating language discrimination when only intonational cues are present. We will achieve this goal by studying language discrimination of Western Catalan and Spanish, two languages primarily differing in some intonational dimensions.

Spanish and Catalan (both Eastern and Western) have been described as syllable-timed languages (Gavaldà-Ferré, 2007; Prieto et al., 2012; Ramus et al., 1999; Ramus, Dupoux, & Mehler, 2003) and there is evidence that at 4–5 months of age, infants can discriminate Spanish from Eastern Catalan (Bosch, Cortés, & Sebastián-Gallés, 2001; Bosch & Sebastian-Galles, 1997, 2001), which differ on the basis of several phonological dimensions (see Table 1 for a summary of some relevant phonological differences between Spanish and Catalan). Spanish has a five-vowel system, Western Catalan has a seven-vowel one and Eastern Catalan an eight-vowel one (Carbonell & Llisterri, 1999; Martínez-Celdrán, Fernández-Planas, & Carrera-Sabaté, 2003).

Table 1
Summary of relevant phonological characteristics of Spanish and Catalan at the segmental and intonational levels.

	Spanish	Catalan
Number of vowels[a,b].(stressed vs. *unstressed* positions)	5 (5 / 5)	8 (Eastern) (7 / 3) 7 (Western) (7 / 7)
Main difference in boundary cues' frequency[c,d]	Pitch reset	Preboundary lengthening
Main nuclear pitch accent[c]	L* + H	L + H*
Foot stress[e]	Trochaic	Iambic
Intonational Phrasing Pattern[f]	Most common (S)(VO)	Most common (S)(VO), Also (SV)(O)

[a] Carbonell and Llisterri (1999).
[b] Martínez-Celdrán et al. (2003).
[c] Frota, D'Imperio, Elordieta, Prieto, and Vigário (2007).
[d] Prieto et al. (2012).
[e] Gibson (2010); Ohannesian (2005); Oliva (1992); Wheeler and Wheeler (2005).
[f] Imperio, Elordieta, Frota, Prieto, and Vigário (2009).

Of the most distinctive phonological features differentiating Spanish and Western Catalan from Eastern Catalan is the existence of vowel reduction in the latter, but not in the former. Languages with vowel reduction have fewer vowels in unstressed positions as compared to stressed positions, while in languages without vowel reduction the same vowels appear in stressed and unstressed positions. For example, central vowels (/e/ and /o/) do not appear in unstressed syllables in Eastern Catalan, but they do in Spanish and Western Catalan (the city of Barcelona is pronounced /bərsəl'onə/, /barθel'ona/, in Eastern Catalan and Spanish, respectively). Bosch and Sebastian-Galles (2001) suggested that the existence of vowel reduction in Eastern Catalan, but not in Spanish, could render the languages sufficiently different in terms of their rhythm to allow their discrimination by 4–5 months of age, as vowel reduction is often found in stress-timed languages (Dauer, 1983; Roach, 1982). However as aforementioned, analyses using metrics (Gavaldà-Ferré, 2007; Prieto et al., 2012; Ramus et al., 1999) and perceptual experiments (Ramus et al., 2003) support the rhythmic similarity between them. Critical to our research proposals, the existence of vowel reduction in Eastern Catalan but not in Spanish and Western Catalan makes the three languages very different in terms of the token (frequency) distribution of vowels in the signal.

Eastern Catalan has relatively few mid vowels in stressed positions (16% Alcina and Blecua (1975) of all the vowels in the language), while there are three times more mid vowels in Spanish (49% Rafel, 1980). The occurrence of mid front vowels also differs greatly in the two Catalan dialects (Wheeler & Wheeler, 2005).Western Catalan does not have vowel reduction, rendering its distribution of vowels closer to Spanish (48% estimate from Fig. 2 from Ortega-Llebaria & Bosch, 2015). Ortega-Llebaria and Bosch (2015) also showed that the frequency of specific vowels (/ə/, /e/, /a/) is a reliable cue to classify an utterance belonging to either Western or Eastern Catalan, a result supporting the existence of differences in vowel distribution between the two Catalan dialects.

Spanish and Catalan also differ at other prosodic levels, in particular intonation. As Prieto et al. (2015) showed that Eastern and Western Catalan do not differ in significant intonational dimensions, we will refer to both dialects as Catalan here. Concerning intonational information, Spanish and Catalan tend to group together when they are compared to other Romance languages such as Italian or European Portuguese, but they still show significant differences in the frequency of intonational boundaries and pitch movement, i.e. rising or falling (see Table 1 for a detailed description, Frota et al., 2007; Hualde, 2003; Prieto et al., 2012). At the foot level, Spanish is clearly trochaic, but Catalan has been described as iambic (Gibson, 2010; Ohannesian, 2005; Oliva, 1992; Wheeler & Wheeler, 2005). The most common phrasing intonational pattern in both Catalan and Spanish is (S)(VO) (Imperio et al., 2009). However, Catalan shows a tendency to divide constituents of an utterance into equally long parts, which yields to a frequent (SV)(O) pattern. This pattern is of very low frequency in Spanish.

Summing up, the three languages have been described as very similar at the rhythmic level, but they cluster in different ways at the vowel distribution and intonation levels. Spanish and Western Catalan present very similar vowel token distributions, while Eastern Catalan differs from them. In terms of intonation, the two Catalan dialects are very similar, Spanish being the distinctive one. We hypothesize that at 4–5-months of age infants may use distributional properties of the vowel sounds of the languages. We also hypothesize that they may use other prosodic cues, i.e. intonation, to discriminate languages.

The necessity to discriminate languages is the primary distinctive feature of early language acquisition in bilingual infants when compared to monolingual ones. However, the number of studies comparing monolingual and bilingual infants in the first six months of life is quite reduced. Such studies have systematically showed that monolingual and bilingual infants can discriminate the same language pairs: Byers-Heinlein et al. (2010), who investigated discrimination of English and Tagalog in newborns, Molnar et al. (2014), who investigated discrimination of Basque and Spanish in 3.5-month-old infants and a series of studies investigating the discrimination of Spanish and (Eastern) Catalan in 4–5 month-old infants (Bosch & Sebastian-Galles, 1997, 2001; Nacar Garcia, Guerrero-Mosquera, Colomer, & Sebastian-Galles, 2018). An additional goal of the present research is to provide further data to this critical process in infants growing up in bilingual environments. As described, Spanish and Eastern Catalan differ in terms of vocalic distribution *and* at the intonational level, thus, in Bosch and Sebastian-Galles (2001), infants had two different types of information at their disposal in order to discriminate languages. In the present experiments, language pairs will differ in just one dimension, potentially making discrimination more difficult. Some studies have shown that bilinguals can visually discriminate languages (with no auditory information, silent videos of talking speakers) whereas monolinguals cannot (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007). These results have led researchers to conclude that bilinguals may be able to perform finer language discriminations than monolinguals. Nevertheless, Weikum et al. (2007) reported equivalent visual language discrimination between monolinguals and bilinguals at 6 months of age; it was at 8 months of age when bilinguals and monolinguals differed. Considering the lack of differences in language discrimination abilities between monolinguals and bilinguals before 6 months of age, we do not expect any difference between the two populations in the present research.

In the present research, we have run 4 different experiments. In Experiment 1 we test infants' ability to discriminate two languages with equivalent prosodic information (rhythm and intonation) but differing mainly in the distribution of vocalic information: Western and Eastern Catalan. Successful discrimination will be a proof of infants' sensitivity to differences in vocalic distributions. In Experiment 2, we investigate the discrimination of Western Catalan and Spanish. As already mentioned, these two languages mainly differ at the intonational level; therefore, discrimination will provide additional support to the hypothesis that intonational information can be a reliable cue for language discrimination at 4–5 months of age. In experiments 1 and 2 the performance of infants growing up in monolingual and bilingual environments will be compared. Experiments 3 and Experiment 4 mirror the first two experiments with low-pass filtered sentences. Low-pass filtering has been extensively used to investigate early language discrimination (Bosch & Sebastian-Galles, 1997; Chong et al., 2018; Johnson et al., 2003; Nazzi et al., 1998). By filtering frequencies above 400 Hz most segmental information is removed while most prosodic information is preserved (i.e. intonation and stress). This manipulation of the speech signal should hinder the discrimination of Western from Eastern Catalan, but Western Catalan from Spanish should still be distinguishable. Experiments 3 and 4 are carried out only with monolingual participants.[2]

2. General methods

Stimuli and procedure followed Bosch and Sebastian-Galles (2001) as close as possible to allow comparison of results.

2.1. Participants

Participants in the four experiments were healthy, full-term infants (>37 gestation weeks) coming from either Catalan and/or Spanish speaking families. Infants were recruited from maternity rooms from private hospitals in Barcelona (Spain). The linguistic background of the participants was assessed using an adapted version of the Bosch and Sebastian-Galles (2001) language questionnaire. In Experiment 1 and Experiment 3 participants were primarily exposed to Eastern Catalan (monolinguals and Eastern-Catalan dominant bilinguals), while in Experiment 2 and Experiment 4, participants were primarily exposed to Spanish (monolinguals and Spanish-dominant bilinguals). Table 2 displays the distribution of participants across experiments. Monolingual participants were exposed at least 88% to the dominant language and bilingual participants were exposed to the dominant language at a level ranging from 51% to 78%. Participants' data were discarded if they were exposed to Western Catalan on a regular basis. This language criterion was imposed on participant selection in all 4 experiments. The research reported in this article was conducted in accordance with the principles expressed in the Declaration of Helsinki and approved by the local ethical committee (the clinical research ethical committee of the Parc de la Salut Mar). Caregivers signed a consent form before participating in the experiment and a small gift was given to them at the end of the experiment.

2.2. Stimuli

The stimuli were sentences in Eastern Catalan, Western Catalan, and Spanish. A female proficient Catalan-Spanish bilingual speaker who was an actress student, highly proficient in both Catalan dialects and Spanish, was recorded. The native speaker had no detectable accent in any of the three languages. Three independent native speakers could not detect the dominant/first language of the speaker. The stimuli were recorded with an Audio-Tecnica microphone (AT2050) at a sampling rate of 44,100 Hz, in a soundproof room at the 'Laboratori de Recerca en Infancia' at University Pompeu Fabra. Audacity® recording and editing software (Mazzoni, 1999) was used to record the stimuli and to extract the sentences from the recordings. The speaker was recorded while she was reciting sentences describing a child's storybook without text, named *Frog, where are you?* by Mayer (1969). She was instructed to recite the sentences in an infant-directed manner, avoiding unnecessary pauses, so as to ensure natural, smooth and continuous recordings.

For each language, a total of 12 sentences were selected. All the sentences were declarative, equivalent to phonological phrases in the prosodic hierarchy. A content word was never repeated more than three times in the whole experiment and the three repetitions never appeared in the same phase of the study, i.e. familiarization or test. The sentences in the different languages were similar in meaning, but not direct translations. The sentences were organized into four different, three-sentence long, passages per language. The order of the sentences within each passage was randomized across participants. We measured duration, number of syllables and different aspects of fundamental frequency (mean, maximum, minimum and standard deviation) using the Praat acoustic analysis software (Boersma & Weenink, 2016), for

Table 2
Description of participants' language background, number of participants in each experiment, language used in the familiarization phase in the four experiments and average exposure to Dominant Language.

Experiment	Dominant lang/ familiarization language	Monolinguals		Bilinguals	
		n	DomL %	n	DomL %
1: Eastern vs Western Catalan	Eastern Catalan	20	97.5	20	63
2: Spanish vs Western Catalan	Spanish	20	97	20	66
3: Low pass East vs West Cat	Eastern Catalan	20	93	–	–
4: Low pass Sp vs West Cat	Spanish	20	90.7	–	–

[2] Lack of availability of bilingual participants motivated their exclusion in experiments 3 and 4.

sentences and passages in each language (see table S4 in the Supplementary Information). No statistical differences between any pair of languages were observed for any measurement at the sentence or passage level (unpaired two sample t-tests, uncorrected p-values>.18). As shown in Fig. 1, the vowels contained in our sentences were representative of the distribution of vowel tokens in the three languages.

Following Bosch and Sebastian-Galles (2001), a trial consisted of one passage, repeated twice, which resulted in six sentences per trial with 1 s silence between them. The average duration of a trial was 28 s. In the familiarization phase, participants always heard sentences in their dominant language, either Eastern Catalan or Spanish. Two passages were used for the familiarization. Participants had to accumulate 2 min paying attention to the screen while listening to the passages. To do so, they had to at least be presented with three trials per passage. The test phase consisted of eight trials, which were organized into two blocks for the purposes of the analyses following Bosch and Sebastian-Galles (2001). Each block contained two trials in the familiarized language and two trials in the non-familiarized language. The stimuli in the test were all novel. Their order and side of presentation was randomized, with the exception that the first trial of each block was in the non-familiarized language.

In the first two experiments, the sentences were natural speech: in Eastern and Western Catalan in Experiment 1 and in Spanish and Western Catalan in Experiment 2. In Experiment 3 and Experiment 4, the same stimuli were used respectively, only now the sentences were low-pass filtered with a cut-off threshold at 400 Hz and a roll-off at 48 dB/octave, which were edited using Audacity®. In the last two experiments, most segmental information was removed, leaving primarily prosodic cues available, whereas in the first two experiments sentences had both segmental and prosodic cues available (see in the Supplementary Information Fig. S2 and S3 an example of the natural and low-pass filtered stimuli).

2.3. Procedure

Infants were tested in a sound-attenuated, dimly lit room at the 'Laboratori de Recerca en Infancia' (Center for Brain and Cognition, Universitat Pompeu Fabra). The testing room was equipped with three ASUS VE276N (size: 27″, resolution: 1920 × 1080) monitors facing a chair. The monitors were positioned one right next to the other and the chair was oriented towards the one in the middle. Each lateral screen subtended a visual angle of 21°, Two M-Audio AV 30 loudspeakers were hidden behind the two lateral monitors. The experiment was run by an experimenter outside the testing room using a custom-made program (based on Matlab and Psychtoolbox (Brainard, 1997)). A SONY HC9 camera was placed 20 cm above the central monitor allowing experimenters to record infants' reactions for online and offline analyses. The camera recorded not only the participant, but also a panel with led infrared light bulbs placed behind the chair (see fig. S4 in Supplementary material for a detailed description) and a mirror mounted on top of the light panel. The panel and mirror were used online as a control for stimuli presentation and for the offline coding. A text file was also generated for each participant with the order of the sentences in each trial and the side of presentation (left or right lateral monitor). The participant was seated on the caregiver's lap at approximately 75 cm from the central monitor. The caregiver listened to masking music during the experiment and was instructed to avoid any interaction with the infant.

We used the adaptation of the familiarization-preference paradigm (Jusczyk et al., 1993), developed by Bosch and Sebastian-Galles (2001) to test language discrimination with this particular age group and population. Each trial began with a rotating flower appearing in the central screen. The rotating flower remained on the screen until the participant fixated on it. Then the central flower was extinguished, and a second flower of different colors appeared in one of the two lateral monitors (see Fig. S1 in Supplementary material). Once the participants fixated on that lateral monitor, sentences were played through the corresponding lateral loudspeaker in the same side. The side of the stimuli presentation was randomized with a maximum of three consecutive trials coming from the same side.

Infants were first familiarized in their native language, either Eastern Catalan (Experiment 1, Experiment 3) or Spanish (Experiment 2, Experiment 4). Once participants had accumulated two minutes of total looking time to the screens during familiarization, the test phase began. The order of the test trials was pseudorandomized, the first test trial in each block was always in the non-familiarized language. No more than

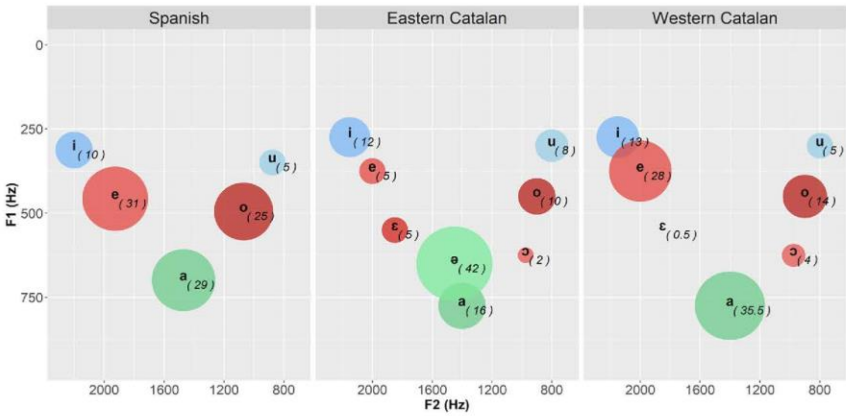


Fig. 1. A visual representation of the differences in vowel distribution in our stimuli, Western Catalan and Spanish being more similar than Eastern Catalan. High vowels are in blue, mid vowels are in red and low-central vowels in green. Percentages of vowel distribution (in parentheses) extracted from our stimuli. Our stimuli closely mimic the vowel distribution as the ones described for Eastern Catalan in Rafel (1980), for Western Catalan in Ortega-Llebaria and Bosch (2015) and for Spanish in Alcina and Blecua (1975). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

two consecutive passages of the same language were played. Trial order was counterbalanced across participants. The whole experiment lasted between 7 and 10 min.

2.4. Coding

An experimenter sat outside the testing room, coded online, while watching via a camera the participants' behavior without sound. If the participant was looking at the screen, then the researcher pressed a key continuously, if they looked away, then the experimenter stopped pressing, which enabled the stimulus presentation to be infant-controlled. Looking times were coded offline using PsyCode (http://psy.cns.sissa.it). Looking time was computed if the participants had turned their heads and fixated on the screen where the visual stimulus was presented. The time participants did not pay attention to the screen, judging from the videos were subtracted from the total looking time. The maximum looking time for a given trial was on average 28 s. Looking times under 500 ms were discarded and were counted as missing data. Coders were blind to the experimental condition since the videos had no sound. Information on the specific language and sentences heard in each trial was saved on a separate text file. The onset and offset of the sentences were evident from the led panel shown in the videos. A second coder analyzed half of the videos for each experiment. We calculated the Interclass Correlation Coefficient (ICC) to check for agreement between the two sets of observations by a two-way model with raters as a fixed-factor and population as a random factor (R 'irr' package (Gamer, Lemon, Fellows, & Singh, 2019; Koo & Li, 2016; Shrout & Fleiss, 1979). ICCs oscillated between 0.915 (Experiment 2) and 0.989 (Experiment 3) showing a high degree of agreement. Reported data were coded by the primary coder.

2.5. Data analysis

Statistical analyses and figures were generated using R and RStudio (R Core Team, 2019; RStudio Team, 2019). We used the 'ez' R package (Lawrence, 2016) to perform the ANOVAs. The eight test trials were divided into two blocks for the purposes of the analyses, following Bosch and Sebastian-Galles (2001). We only included participants with at least one valid data point per block and condition. Percentage of missing data-points were 0%, 0%, 1.2% and 1.8% for experiments 1–4 respectively. We also conducted paired *t*-tests and Bayesian Factor analysis to compare the looking times in each Condition (switch vs. same) using JASP (van Doorn et al., 2019). In experiment 1,2 and 4 we calculated the Bayes Factor in favor of the alternative hypothesis (BF_{10}), i.e. we expect participants to be able to discriminate. In experiment 3 we calculated the Bayes Factor in favor of the null hypothesis (BF_{01}) since based on our hypothesis we expect infants to be unable to discriminate Western and Eastern Catalan when low-pass filtered.

3. Experiment 1 testing discrimination based on vowel distribution

In Experiment 1, we tested the hypothesis that 4–5 month-old infants can discriminate languages when primarily differences in the distribution of vocalic information are present. We presented infants with sentences of two dialects of Catalan (Eastern and Western) differing in the distribution of vocalic sounds and with equivalent rhythmic and intonational information. Infants raised in monolingual and bilingual environments were also compared.

3.1. Methods

The final sample consisted of 20 monolinguals (10 females, 10 males, M_{age} = 150 days, SD = 7.2, range 138–163, 97,5% exposure to Eastern Catalan) and 20 bilinguals (7 females, 13 males, M_{age} = 148 days, SD = 8.24, range 136–161, 63% exposure to Eastern Catalan). The monolingual participants were exposed to Eastern Catalan at least 85% of the time and the bilingual participants were exposed to Eastern Catalan and Spanish with a maximum exposure to Eastern Catalan 78% of the time. One of the bilingual participants had a 2% exposure to English from a relative. An additional 15 infants were tested but not included in the final sample due to: fussiness and/or lack of interest (7), crying (6), experimental error (2). Nineteen infants were also tested but not included in the final sample because they failed to reach the inclusion criteria: language profile (13), minimum weight at birth ($\leq$ 2600 g) (3) or health issues (3).

Infants were familiarized in their dominant language, Eastern Catalan. During the testing phase, they were presented with sentences in the same language as the familiarization (same condition) and sentences from a different language, here Western Catalan (switch condition).

3.2. Results and discussion

Total looking times to the Same, i.e. Eastern Catalan, and Switch, i.e. Western Catalan, language trials during the test phase are shown in Fig. 2. A three-factor mixed ANOVA was performed with Condition (Same, Switch) and Block (1,2) as within-subjects factors and Language as a between-subjects factor (monolingual, bilingual). A significant effect of Condition ($F(1, 38) = 14.21$, $p < .001$, $\eta^2_G = 0.03$) and Block ($F(1, 38) = 30.58$, $p < .001$, $\eta^2_G = 0.10$) were found. Participants looked longer in the Switch condition (M = 14,887 ms, SD = 5721) than in the Same condition (M = 13,032 ms, SD = 5909). They also looked longer in the first Block (M = 15,796, SD = 5674) than in the second Block (M = 12,123, SD = 5511). The interaction between the two did not reach statistical significance ($F(1, 38) = 0.10$, $p = .759$, $\eta^2_G < 0.01$). There was no effect of Language ($F(1, 38) = 0.48$, $p = .495$, $\eta^2_G < 0.01$), nor did it interact with Condition ($F(1, 38) = 1.09$, $p = .303$, $\eta^2_G < 0.01$) nor with Block ($F(1, 38) = 1.47$, $p = .233$, $\eta^2_G < 0.01$). The triple interaction between the three factors was also not significant ($F(1, 38) = 0.31$, $p = .580$, $\eta^2_G < 0.01$). Separate paired t-tests for monolinguals and bilinguals on the Condition factor showed significant differences for each group: $t(19) = 2.20$, $p = .041$, $d = 0.49$, $BF_{10} = 1.63$ and $t(19) = 3.07$, $p = .006$, $d = 0.69$, $BF_{10} = 7.50$; for monolinguals and bilinguals, respectively. When we consider the data of both monolinguals and bilinguals together, we have a very strong effect of discrimination ($BF_{10} = 52.16$). We found that 4.5-month-old infants were able to discriminate Eastern from Western Catalan regardless of their linguistic background. As the main difference between the two Catalan dialects is their vowel distribution, the results support the hypothesis that some segmental information, in particular, vowel distribution may be used as a cue for language differentiation before the vowel repertoire of the native language is established.

4. Experiment 2 testing discrimination based on suprasegmental information

Experiment 2 was designed to test the hypothesis that 4–5 month-old infants can rely on intonational information to discriminate languages. Previous research has shown that five-month-old infants can discriminate between Eastern Catalan and Spanish (Bosch & Sebastian-Galles, 1997, 2001). However, Eastern Catalan and Spanish differ at the intonational level and also at the vowel distribution. In Experiment 2, we attempted to extend the results by comparing a new pair of languages, Western Catalan and Spanish that primarily differ at the intonational level.

4.1. Methods

The final sample consisted of 20 monolinguals (10 females, 10 males, M_{age} = 150 days, SD = 7.15, range 138–161, 97% exposure to Spanish) and 20 bilinguals (10 females, 10 males, M_{age} = 148 days, sd = 7.94, range 137–164, 66% exposure to Spanish). The monolingual

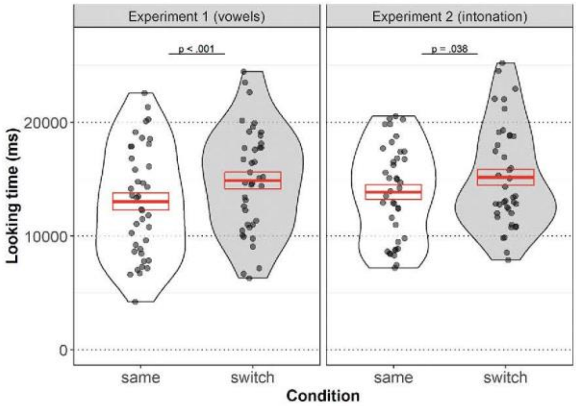


Fig. 2. Average looking times to trials in the same and switch condition in Experiment 1 testing vowel distribution (Eastern Catalan vs. Western Catalan with natural stimuli), and in Experiment 2 testing intonational information (Spanish vs. Western Catalan with natural stimuli). Each dot represents the average of one participant. The boxes and whiskers represent the mean and the standard error, respectively.

participants were exposed to Spanish at least 87% of the time and the bilingual participants were exposed to Spanish and Eastern Catalan with a maximum exposure to Spanish 78% of the time. Two of the bilingual infants had each a 5% and a 6% exposure to English through a relative. An additional 11 infants were tested but not included in the final sample due to: fussiness and/or lack of interest (1), crying (6), technical error (4). Sixteen infants were also tested but not included in the final sample because they failed to reach the inclusion criteria: language profile (10), minimum weight at birth ($\leq$ 2600 g) (2), health issues (4).

Infants were familiarized in their dominant language, Spanish. During the testing phase, they were presented with sentences in the same language as in the familiarization phase (same condition) and with sentences from a different language, here Western Catalan (switch condition).

4.2. Results and discussion

Total looking times to the Same, i.e. Spanish, and Switch, i.e. Western Catalan, language trials during the test phase are shown in Fig. 2. A three-factor mixed ANOVA was performed with Condition (Same, Switch) and Block (1, 2) as within-subjects factors and Language as a between-subjects factor (monolingual, bilingual). A significant effect of Condition ($F(1, 38) = 4.62$, $p = .038$, $\eta^2_G = 0.02$) and Block ($F(1, 38) = 24.60$, $p < .001$, $\eta^2_G = 0.12$) were found. Participants looked longer in the Switch condition (M = 15,163 ms, SD = 5597) than in the Same condition (M = 13,863 ms, SD = 5547). They also looked longer in the first Block (M = 16,381, SD = 5009) than in the second Block (M = 12,645, SD = 5549). The interaction between the two did not reach statistical significance ($F(1, 38) = 0.02$, $p = .878$, $\eta^2_G < 0.01$). There was no effect of Language ($F(1, 38) = 0.71$, $p = .406$, $\eta^2_G < 0.01$) nor did it interact with Condition ($F(1, 38) = 0.00$, $p = .979$, $\eta^2_G < 0.01$) nor with Block ($F(1, 38) = 1.52$, $p = .225$, $\eta^2_G < 0.01$). The triple interaction between the three factors was also not significant ($F(1, 38) = 0.01$, $p = .903$, $\eta^2_G < 0.01$). Separate paired *t*-tests for monolinguals and bilinguals on the Condition factor failed to show significant differences for each group: $t(19) = 1.54$, $p = .141$, $d = 0.34$, $BF_{10} = 0.63$ and $t(19) = 1.51$, $p = .149$, $d = 0.34$, $BF_{10} = 0.61$; for monolinguals and bilinguals, respectively. When we consider the data of both monolinguals and bilinguals together, we have a small effect of discrimination ($BF_{10} = 1.41$).

We found that 4.5-month-old infants could discriminate Western Catalan from Spanish in a similar way regardless of their linguistic background. Because these two languages primarily differ at the intonational level, the present results extend previous studies showing the use of intonational information to discriminate languages at 4–5 months of age.

Concerning language background, given the lack of effects of this variable (see Table 3), it was no longer checked in the following experiments, that included only monolingual infants.

In Experiment 1 and Experiment 2, we showed that 4–5-month-old infants can discriminate between languages differing mainly in distributional properties of their vocalic system and in intonational information. Experiment 3 and Experiment 4 were designed to further test these hypotheses. We low-pass filtered the stimuli used in Experiment 1 and Experiment 2 and tested two new sets of participants. Low-pass filtering has been widely used as a way of removing most of the segmental information and leaving prosody intact (Bosch & Sebastian-Galles, 1997; Chong et al., 2018; Molnar et al., 2014; Thierry Nazzi et al., 1998). If infants used the vowel distribution in Experiment 1 as a cue for discrimination, we do not expect infants to discriminate Low-pass filtered Western from Low-pass filtered Eastern Catalan (Experiment 3). In contrast, in Experiment 4 we expect infants to discriminate low-pass filtered Spanish from low pass filtered Western Catalan, as low pass filtering will not have removed the intonational information that allowed discrimination in Experiment 2.

5. Experiment 3 testing discrimination with low-pass-filtered stimuli to control for differences in vowel distribution

In Experiment 3, we investigated whether suprasegmental

Table 3
Mean looking times and standard deviations to the two Conditions for Monolingual and Bilingual participants in experiments 1 and 2.

Experiment	Monolinguals		Bilinguals	
	Same	Switch	Same	Switch
1: West vs. East Catalan	12,805 (SD = 4466)	14,145 (SD = 5382)	13,258 (SD = 5200)	15,627 (SD = 3534)
2: West Cat vs. Spanish	13,357 (SD = 4746)	14,641 (SD = 4874)	14,368 (SD = 3374)	15,684 (SD = 4095)

information is enough for 4.5-month-old infants to discriminate between Western and Eastern Catalan. By low-pass filtering the stimuli used in Experiment 1, we eliminated most of the segmental information. Pertinent to our goals, we significantly attenuated the information on the differences in vowel distribution. If infants fail to discriminate the two, it would provide supportive evidence to our hypothesis that infants used the frequency of vowels of the two Catalan dialects as a cue to discriminate in Experiment 1.

5.1. Methods

The final sample consisted of 20 infants (14 females, 6 males, M_{age} = 149 days, SD = 10.11, range 139–168, mean exposure to Eastern Catalan 93%). An additional 10 infants were tested but not included in the final sample due to: fussiness and/or lack of interest (3), crying (4), technical error (2), parental interference (1). Seven infants were also tested but not included in the final sample because they failed to reach the inclusion criteria: language profile (5), health issues (2).

Infants were familiarized and tested in the same way as in experiment 1, only now the stimuli were low-pass filtered at 400 Hz.

5.2. Results and discussion

Total looking times to the Same, i.e. Eastern Catalan, and Switch, i.e. Western Catalan, trials during the test phase are shown in Fig. 3. A two-factor repeated measures ANOVA was conducted with Condition (switch, same) and Block (1, 2) as factors. Condition ($F(1, 19) = 0.00$, $p = .957$, $\eta^2_G < 0.01$) did not reach statistical significance, neither did Block ($F(1, 19) = 1.03$, $p = .323$, $\eta^2_G < 0.01$). The interaction was also not significant ($F(1, 19) = 0.69$, $p = .417$, $\eta^2_G < 0.01$). In order to make sure that our null effect of Condition was not due to lack of power, we calculated the Bayes Factor for the null hypothesis against the alternative one, i.e. infants were able to discriminate. The result suggests that the null model is 4.29 (BF_{01}) times more probable than the alternative model, which is a moderate effect. Therefore, we have evidence of absence of an effect (Dienes, 2014).

Participants looked equally in the Switch condition (M = 11,386 ms, SD = 5835) and in the Same condition (M = 11,329 ms, SD = 5720). The lack of discrimination speaks to the nature of the cue that infants used to distinguish between Eastern from Western Catalan in Experiment 1, supporting the conclusion that infants used differences in vowel distribution, as discrimination was not possible when this information was practically removed from the stimuli. Our findings also support the hypothesis that the two dialects do not differ in terms of intonational contours, as low-pass filtering does not remove this type of prosodic features.

6. Experiment 4 testing discrimination with low pass filtered stimuli to check for suprasegmental information sufficiency

In Experiment 4, we tested whether intonational cues are sufficient for 4.5-month-old infants to discriminate between Western Catalan and Spanish. We low-pass filtered the stimuli used in Experiment 2, in order to leave behind only suprasegmental information. If our hypothesis that infants can use the differences in intonational patterns to discriminate between the two is correct, then they should succeed even when no segmental information is available.

6.1. Methods

The final sample consisted of 20 infants (10 females, 10 males, M_{age} = 151 days, SD = 7.33, range 138–164, mean exposure to Spanish 90.7%). One of the participants had an exposure of 2% to English by a relative. An additional 6 infants were tested but not included in the final sample due to: fussiness and/or lack of interest (2), cried (2), technical error (1), parental interference (1). One infant was also tested but not included in the final sample because they failed to reach the inclusion criteria (health issues).

Infants were familiarized and tested in the same way as in Experiment 2, only now the stimuli were low-pass filtered at 400 Hz.

6.2. Results and discussion

Total looking times to the same, i.e. Spanish, and switch, i.e. Western Catalan, language trials during the test phase are shown in Fig. 3. A two-factor repeated measures ANOVA was conducted with Condition (switch, same) and Block (1, 2) as factors. Only Block reached significance ($F(1, 19) = 20.74$, $p < .001$, $\eta^2_G = 0.09$) and Condition was close to it ($F(1, 19) = 3.82$, $p = .066$, $\eta^2_G = 0.01$). The interaction between Condition and Block ($F(1, 19) = 2.58$, $p = .124$, $\eta^2_G < 0.01$) was not

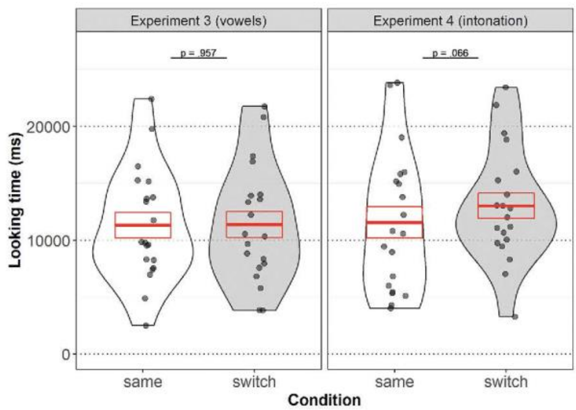


Fig. 3. Average looking times to trials in the same and switch condition in exp.3 Eastern Catalan vs. Western Catalan with low-pass filtered stimuli, and in exp.4 Spanish vs. Western Catalan with low-pass filtered stimuli. Each dot stands for the average of one participant. The boxes and whiskers for the mean and the standard error.

significant. Participants showed a trend to look more in the Switch condition (M = 13,036 ms, SD = 6055) than in the Same condition (M = 11,575 ms, SD = 6667).

Infants showed a trend to discriminate between low-pass filtered Spanish and Western Catalan (BF_{10} = 1.125). The lack of significant differences does not allow us to draw any firm conclusions on whether infants can discriminate the two languages when low-pass filtered.

7. General discussion

In a series of four experiments, we have investigated 4–5-month-old infants' abilities to discriminate languages using information other than linguistic rhythm. By comparing language discrimination in pairs of languages, differing in their vocalic distribution or in their intonational patterns, we have been able to assess the relevance of these factors in early language differentiation. In Experiment 1 and Experiment 3 we tested infants' capacities to discriminate two languages differing primarily in the distribution of vocalic information (Eastern Catalan and Western Catalan). Infants successfully discriminated the two languages only when segmental information was available (Experiment 1). In Experiment 3 we low-pass filtered the sentences and discrimination was no longer observed. In Experiment 2 and Experiment 4 we observed that 4–5 month-old infants can discriminate languages differing primarily in terms of their intonational patterns (Western Catalan and Spanish), even when low pass-filtered. The comparison of monolingual and bilingual infants in Experiment 1 and Experiment 2 showed no differences in the language discrimination abilities between the two populations.

The primary goal of our investigation was to assess if before infants have established the vowel categories of their native language, they are sensitive to differences in distributional properties of vocalic information in two languages. The results have provided compelling evidence that they are. The results challenge the assumption that vocalic information does not contribute, in a significant way, to early language discrimination abilities. Two types of experimental evidence have supported such an assumption. First, language discrimination studies have shown that on the one hand, patterns of discrimination do not change when segmental information is cancelled by either low-pass filtering or by resynthesizing speech; on the other hand, discrimination is hindered when prosodic information is altered or cancelled. Such results have led researchers to conclude that vocalic (segmental) information plays a negligible role. Second, before 6 months of age infants do not show evidence of perceptual narrowing to the phoneme system of their native language. The earliest narrowing has been reported at around 6 months of age for vowels (Kuhl, 1991; Kuhl et al., 1992; Polka & Werker, 1994; Werker & Tees, 1984). Investigating the same population tested here, Bosch and Sebastián-Gallés (2003) showed that Spanish and Catalan 4–5 month-old monolingual infants were able to differentiate the Catalan-specific /e/–/ɛ/ vowel contrast, a contrast that at 8 months of age Spanish monolingual infants did no longer differentiate (see Albareda-Castellot, Pons, and Sebastián-Gallés (2011) for converging results with a different paradigm). The lack of tuning to the native vocalic repertoire also supported the idea that at this early age, language-specific segmental information plays little role, if any, in early language discrimination abilities. Studies that have considered cross-language differences at the segmental level as a potential cue, have compared the size or overlap of vowel or consonant repertoires. For instance, German possesses front rounded vowels and English does not (Chong et al., 2018). However, considering differences or similarities in the phoneme inventories assumes that perceivers already possess phoneme categories. Our proposal is that infants compute distributions of tokens in the acoustic/perceptual space, without necessarily classifying tokens as members of specific phonetic categories. If two vocalic distributions are different enough, discrimination will be possible.

Sensitivity to vocalic information at such an early age finds support in different types of evidence. Bouchon et al. (2015) showed that at 5 months infants already notice vowel mispronunciations of their name, but not consonant mispronunciations. The first signs of perceptual narrowing occur earlier for vowels, i.e. at 6 months, than for consonants, i.e. at 10 months (Kuhl et al., 1992; Werker & Tees, 1984). Together with acoustic saliency, frequency of occurrence may facilitate such early perceptual narrowing. Across languages of the world, there are less vowel categories than consonant categories. According to Maddieson (2013), the C/VQration (resulting from dividing the number of consonants (C) by the number of vowel qualities (VQ) in a language) there are no languages with more, or even equal, consonants than vowels. For the languages tested in the present research, such number is around 3.8 for Spanish and 2.6 for Catalan (with slight variations depending on the dialectal variation), that is there are 3.8 and 2.6 more consonants than vowels in Spanish and Catalan respectively. Yet, vowels represent around 50% of the sounds in the speech signal (according to Ramus et al. (1999) 43.8% in Spanish and 46.6% in Eastern Catalan, no data provided for Western Catalan).

There are two different (though non-mutually exclusive) types of distributions that infants may have used to discriminate Western Catalan from Eastern Catalan in Experiment 1. One possibility is that they have computed the distribution during the familiarization and test phases, and they have used such information to notice the differences between the sentences of the two languages presented in the test phase. There is abundant evidence showing that infants can rapidly compute distributional information of different linguistic units with short experimental exposure (subtle phonetic cues Maye et al. (2002); syllables Saffran, Aslin, and Newport (1996); see Aslin and Newport (2014) for a review). Although Maye et al. (2002) participants were 6 month-olds, the authors suggested that even younger infants could be sensitive to distributional information due to the universality of the statistical learning mechanism (present even across species, see for a review Santolin and Saffran (2018)). It is therefore possible that infants may have used the distributional information provided during the experiment to perform the task.

The second possibility is that infants may already possess a primitive knowledge about the distribution of sounds in their native language. As infants grow older and attain more experience with their language(s), they build a coarse representation of the quality and quantity of the vowel system and they primarily use this information to discriminate the two languages. The pattern of results of previous research is consistent with the use of information corresponding to infants' native language in language discrimination tasks, as discrimination is more likely to be observed if one of the languages is the native one. For instance, Nazzi et al. (2000) did not observe discrimination in 5-month-old infants exposed to American English when listening to Dutch and German sentences. In the same line, Butler et al. (2011) reported that 5-month-olds learning Southwestern British English could discriminate their native dialect from Welsh English, but they failed to do so between the two unfamiliar Welsh English and Scottish English. The fact that in our study infants were familiarized with sentences in their native language prevents us from assessing which of the two types of distributional information was responsible for infants' behavior. One future line of research is testing infants learning languages with completely different vocalic distribution to the ones tested in the experiment. An alternative path is to test infants using preference tests, without familiarization. Such studies are currently under way in our laboratory.

Our investigation also informs about early sensitivity to intonational differences that infants can use to discriminate languages. Investigation into early perception of prosodic information has been dominated by research on the perception of language rhythm, with few studies targeting the relevance of intonation in early language discrimination. Ramus (2002) tested if newborns could perform a between-class discrimination, i.e. Dutch versus Japanese, with different types of information available. Neonates could discriminate the two languages when the stimuli were resynthesized leaving rhythm and some broad phonotactics available. Ramus superimposed common foreign intonational contours on the original Dutch and Japanese sentences. In this case, discrimination was not significant anymore. To our knowledge only Chong et al. (2018) performed an investigation comparable to ours testing 5-month-olds in a within-class discrimination situation. Chong et al. (2018) did not find discrimination of American-English and German in 5-month-olds learning American English. In our research we have found some evidence of discrimination with two prosodically similar languages. On top of potential differences in the types of intonational information differing between the two sets of languages, it is possible that methodological differences between the two studies may also account for the different patterns. In particular, differences in the stimuli may have facilitated discrimination in our case (or hindered in Chong et al. (2018)). Chong et al. (2018) used different speakers in their stimuli (four female native speakers for each language) and we have used only one speaker. Because we wanted to compare our results with Bosch and Sebastian-Galles (2001) we decided to keep the experiments methodologically as close as possible. Because, the evidence of discrimination in our research is not very robust it is difficult to draw firm conclusions from our research concerning the use of intonational information in early language discrimination.

One potential caveat in our research is the use of low-pass filtering (in Experiments 3 and 4). This technique has often been used in similar procedures as a way to eliminate most of the segmental information in the speech signal (e.g. Bosch & Sebastian-Galles, 1997; Chong et al., 2018; Molnar et al., 2014; Nazzi et al., 1998). The use of low-pass filtering has been criticised on the grounds that this manipulation does not remove all segmental information or that it distorts it to some degree. Such criticism may be relevant given that low-pass filtering preserves some rough information about vocalic information. Although we cannot discard that some vocalic information went through the low-pass filtered, it was not enough to allow language discrimination, as infants were not able to discriminate Eastern Catalan from Western Catalan in Experiment 3.

In Experiments 1 and 2 we compared monolingual and bilingual infants in their capacity to discriminate languages. We did not observe any difference between 4.5 month-old- monolingual and bilingual infants. This result is in agreement with previous research in the field. Using similar procedures based on procedures measuring recovery of attention after habituation/familiarization exposures Bosch and Sebastian-Galles (2001), Byers-Heinlein et al. (2010) with newborns, Molnar et al. (2014) with Basque or Weikum et al. (2007) with visual language have systematically failed to report reliable differences between monolinguals and bilinguals. This result has been found with auditory and visual only stimulus, regardless of the typological or rhythmic distance between the languages (Spanish-Catalan; Tagalog-English; Basque-Spanish; French-English). It has often been argued that the need to discriminate languages in bilingual infants would be on the basis of some processing differences between them, either with linguistic or non-linguistic stimuli. In the investigation reported here, language discrimination was quite challenging, as language pairs mainly differed in either vowel distribution or in subtle intonational cues. If bilingual infants were better at language discrimination, the language pair we have used provided one of the best scenarios for them to display their abilities. Other studies not based on habituation-familiarization procedures have uncovered different ways to discriminate languages at 4–5 months of age between monolinguals and bilinguals (Bosch & Sebastian-Galles, 1997; Nacar Garcia et al., 2018), but the existing data does not allow to adjudicate enhanced processing abilities in bilinguals (different, but not necessarily better or worse).

Research on language processing before six months of age has primarily focused on the developmental changes of language discrimination abilities. Linguistic rhythm has been investigated thoroughly, however a model of the interplay of the different types of linguistic information available seems to be lacking. To our knowledge, the only computational model on language discrimination for infants has been put forward by Carbajal, Radek, and Dupoux (2010). The model was made to resemble infants' discrimination results. After brief exposure to French, the model was only able to perform between-class discrimination (French vs. English) and failed to perform within-class discrimination (Spanish vs. Catalan). Our results show that to understand how language discrimination abilities evolve in the first months of life information about distributional vocalic information must be included.

Our proposal supports and extends Jacques Mehler and coworkers TIGRE proposal (In Mehler et al.'s (1996) words: "TIGRE is a gridlike representation of vocalic nuclei in the speech signal" (p. 113). The grid would primarily represent sequential and timing information "each vowel receives an index indicating its duration and amplitude" (p. 113). TIGRE was not designed to account for within-class language discrimination, as it was conceived "as a first-order filtering device that sorts utterances into their adequate [rhythmic] classes" (p. 113). Mehler et al. (1996) declared that TIGRE could not account for within-class language discrimination in the first months of life. Our research complements Mehler et al. (1996) providing evidence that infants do not only represent timing information, but that already in the first months of life some representation of the distribution of vocalic sounds in the speech signal is established and that infants can use it to discriminate languages.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104628.

Declaration of interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by grants from the Spanish Ministerio de Ciencia, Innovación y Universidades (PGC2018-101831-B-I00), and the Catalan Government (SGR 2017-268 and FI-9015-456763; ICREA [Catalan Institution for Research and Advanced Studies] Academia 2019 award). Konstantina Zacharaki was supported by a fellowship of the Catalan Government (FI2018). We would like to thank Judit Gervain, the editor of this special issue, for her poignant remarks throughout the review process. We thank Alba Portet and Katia Pistrin, who worked as secondary coders in Experiment 1 and 2 respectively. We thank Koralia Georgiadi who proofread the manuscript. We also thank Xavier Mayoral and Silvia Blanch for their technical support. We thank Maria Cinta Bertomeu Madueño, for her important implication in the project in its initial stages and Alba Ayneto Gimeno, who helped us in the stimuli creation. We also thank the clinics Quirón and Sagrada Familia that enabled us to recruit participants in their premises. We also thank all families and infants who participated in the experiments.

References

Abercrombie, D. (1982). Elements of general phonetics. Retrieved from https://books.google.es/books?id=bmQhAQAAMAAJ.

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science, 14*(2), 395–401. https://doi.org/10.1111/j.1467-7687.2010.00989.x.

Alcina, F. J., & Blecua, J. M. (1975). *Gramática española* (p. 1991). Barcelona: Ariel.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: definition and development. In *Child Phonology*. https://doi.org/10.1016/b978-0-12-770601-6.50017-6.

Aslin, R. N., & Newport, E. L. (2014). Distributional language learning: Mechanisms and models of ategory formation. *Language Learning, 64*(Suppl. 2), 86–105. https://doi.org/10.1111/lang.12074.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me. *Psychological Science, 16*(4), 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x.

Bosch, L., & Sebastian-Galles, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*(1), 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1.

Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*, 29–49.

Bosch, L., Cortés, C., & Sebastián-Gallés, N. (2001). Early native-language recognition capacities: A study based on male-voices. *Infancia y Aprendizaje, 24*(2), 197–213. https://doi.org/10.1174/021037001316920735.

Bosch, Laura, & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Language and Speech, 46(2–3), 217–243. doi:https://doi.org/10.1177/00238309030460020801.

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science, 18*(4), 587–598. https://doi.org/10.1111/desc.12242.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. https://doi.org/10.1163/156856897X00357.

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy, 16*(4), 392–417. https://doi.org/10.1111/j.1532-7078.2010.00050.x.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science, 21*(3), 343–348. https://doi.org/10.1177/0956797609360758.

Carbajal, M. J., Radek, F., & Dupoux, E. (2010). *Modeling language discrimination in infants using i-vector representations. 1* pp. 889–894).

Carbonell, J. F., & Llisterri, J. (1999). *CARBONELL, J. F.- LLISTERRI, J. (1999) "Catalan", in* (pp. 61–65).

Chong, A., Vicenik, C., & Sundara, M. (2018). Intonation plays a role in language discrimination by infants. In , *23. Infancy*. https://doi.org/10.1111/infa.12257.

Dauer, R. M. (1983). *Stress-timing and syllable-timing reanalyzed. Journal of Phonetics, Vol. 11, pp. 51–62.* Netherlands: Elsevier Science.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*(July), 1–17. https://doi.org/10.3389/fpsyg.2014.00781.

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., … Wagenmakers, E.-J. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. *PsyArxiv*, 31. https://doi.org/10.31234/osf.io/yqxfr. Preprint, (February).

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303.

Fletcher, J. (2010). *Part IV linguistic phonetics the prosody of speech: Timing and rhythm.* Retrieved from. https://doi.org/10.1002/9781444317251.ch15.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics, 54*(3), 287–295. https://doi.org/10.3758/BF03205263.

Frota, S., D'Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. (2007). *The phonetics and phonology of intonational phrasing in Romance* (pp. 131–153). https://doi.org/10.1075/cilt.282.10fro.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). Irr: Various coefficients of Interrater reliability and agreement. Retrieved from https://cran.r-project.org/package=irr.

Gavaldà-Ferré, N. (2007). *Vowel reduction and Catalan speech rhythm. (September)*.

Gibson, M. A. (2010). *Fonología del español: enfoque desde la teoría de.* Universidad Complutense de Madrid.

Hualde, J. I. (2003). El modelo métrico y autosegmental. *Teorías de La Entonación*, 155–184.

Imperio, M. D., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2009). *Intonational phrasing in Romance: The role of syntactic and prosodic structure. (Selkirk 2000).* https://doi.org/10.1515/9783110197587.1.59.

Johnson, E. K., Jusczyk, P. W., & Ramus, F. (2003). The role of segmental information in language discrimination by English-learning 5-month-olds. In *Jusczyk Lab Final Report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93–107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, R. (1992). Linguistic experience alters phnetic perception in infants by 6 months of age. *Science, 255*(5), 606–608.

Ladefoged, P. (1975). A course in phonetics. Retrieved from https://books.google.es/books?id=w_5YAAAAMAAJ.

Lloyd James, A. (1940). *Speech signals in telephony.* London: Sir I. Pitman & Sons.

Maddieson, I. (2013). Consonant inventories. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Retrieved from https://wals.info/chapter/1.

Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and Hearing, 2*(74). Work in Progress.

Maidment, J. A. (1983). Language recognition and prosody: Further evidence. *Speech, Hearing and Language, 1*, 133. Work in Progress.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). INFANTS' recognition of the sound patterns of their own names. *Psychological Science, 6*(5), 314–317. https://doi.org/10.1111/j.1467-9280.1995.tb00517.x.

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association, 33*(2), 255–259. https://doi.org/10.1017/S0025100303001373.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. https://doi.org/10.1016/s0010-0277(01)00157-3.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Books for Young Readers.

Mazzoni, D. (1999). *Audacity®*.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 101–116).

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2.

Mersad, K., & Nazzi, T. (2012). When mommy Comes to the Rescue of Statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development, 8*(3), 303–315. https://doi.org/10.1080/15475441.2011.609106.

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy, 19*(3), 326–337. https://doi.org/10.1111/infa.12041.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U.

Nacar Garcia, L., Guerrero-Mosquera, C., Colomer, M., & Sebastian-Galles, N. (2018). Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. *Scientific Reports, 8*(1), 2770. https://doi.org/10.1038/s41598-018-20824-0.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1–19.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 756–766. https://doi.org/10.1037/0096-1523.24.3.756.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*(1), 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In *The Blackwell Companion to Phonology* (pp. 1–13). https://doi.org/10.1002/9781444335262.wbctp0048.

Nolan, F., & Jeon, H. S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society, B: Biological Sciences, 369*(1658). https://doi.org/10.1098/rstb.2013.0396.

Ohala, J. J., & Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody. *Problèmes de Prosodie: Expérimentations, Modèles et Fonctions, 2*, 123.

Ohannesian, M. (2005). *La asignación del acento en castellano.* Universitat Autònoma de Barcelona.

Oliva, S. (1992). *La mètrica i el ritme de la prosa* (Quaderns Crema).

Ortega-Llebaria, M., & Bosch, L. (2015). Cues to dialectal discrimination in early infancy. *The Phonetics–Phonology Interface: Representations and Methodologies, 335*, 55.

Pike, K. L. (1945). The intonation of American English. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=&AN=&site=eds-live.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421.

Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., … Sichel-Bazin, R. (2015). Intonational phonology of Catalan and its dialectal varieties. In *Intonation in Romance*. https://doi.org/10.1017/CBO9781107415324.004.

Prieto, P., Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication, 54*(6), 681–702. https://doi.org/10.1016/j.specom.2011.12.001.

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans XXIII, 2*, 473–496.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th International Congress of Phonetic Sciences 2003* (pp. 337–342).

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349–351.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. In , *2. Annu. Rev. Lang. Acquis.* https://doi.org/10.1075/arla.2.05ram.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America, 105*(1), 512–521. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9921674.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X.

Roach, P. (1982). *On the distinction between 'stress-timed' and 'syllable-timed' languages.*

RStudio Team. (2019). RStudio: Integrated development environment for R RStudio, Inc. Retrieved from http://www.rstudio.com/.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Santolin, C., & Saffran, J. R. (2018). Constraints on statistical learning across species. *Trends in Cognitive Sciences, 22*(1), 52–63. https://doi.org/10.1016/j.tics.2017.10.003.

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994–999. https://doi.org/10.1177/0956797612436817.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127.

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy, 17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x.

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2005). Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(1), 95–100. https://doi.org/10.1037/0097-7403.31.1.95.

Vicenik, C. (2011). *The role of intonation in language*. University of California at Los Angeles.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159. https://doi.org/10.1126/science.1137686.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Wheeler, M. W., & Wheeler, R. L. M. W. (2005). The phonology of Catalan. Retrieved from https://books.google.gr/books?id=PkoVDAAAQBAJ.

White, L., Floccia, C., Goslin, J., & Butler, J. (2014). Utterance-final lengthening is predictive of infants' discrimination of english accents. *Language Learning, 64* (SUPPL.2), 27–44. https://doi.org/10.1111/lang.12060.

Lawrence, Michael A. (2016). Package "ez". *R package version, 4(0)*.

3. The contribution of intonation in Adult Language Discrimination

Konstantina Zacharaki[*], Nuria Sebastian-Galles[*]

[*] Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Barcelona, Spain

3.1. Abstract

Humans can use a variety of segmental and prosodic cues to discriminate between languages. Previous research has focused on rhythmical cues. Our goal here is to test whether adults can take advantage of intonational cues in order to discriminate low-pass filtered sentences in Catalan and Spanish. We are also interested in whether previous knowledge of the tested languages influences performance, so we tested two groups of participants, one being native and the other one non-native. We found that natives and non-natives were able to discriminate and classify the sentences correctly with just intonational information; however, native participants outperformed non-native participants. We extend previous findings on the relevance of intonational cues and on previous knowledge in adult language discrimination.

3.2. Introduction

The question of how people distinguish one language from another is becoming more important as there is a growing number of individuals living in multilingual environments. Understanding how humans discriminate between languages may provide pivotal knowledge in developing multilingual speech-recognition devices. In the present research we investigate how human adults can discriminate between two prosodically similar languages, when only intonation information is available.

A great amount of research on language discrimination and language identification (LID) has been conducted over the last 50 years (for a review on LID see Komatsu, 2007). Adults tend to rely heavily on phonemic knowledge, since they have in place a fully functioning phonetic repertoire, their native one. Once they detect sounds or combinations of sounds that are not found in their native language, they consider the speech segment as foreign. Previous research has used a variety of different manipulations of the signal to remove segmental information to investigate whether prosodic information is also sufficient for language discrimination. Such techniques are low-pass filtering (Atkinson, 1968; Bush, 1967), speech resynthesis (Ramus & Mehler, 1999), laryngographic signal (Maidment, 1976, 1983; Moftah & Roach, 1988), pulse train signal (Ohala & Gilbert, 1979) and Linear Prediction coefficients (LPC) filtering (Komatsu et al., 2002). However, these techniques do not disentangle different types of prosodic information, i.e. rhythm and intonation.

Languages have been classified into different categories based on their prosodic properties, in particular, rhythm. Lloyd James (1940) was the first to notice differences in languages' rhythms, which led him to coin the terms 'machine-gun rhythm' for languages such as Spanish or Italian and 'Morse code rhythm' for languages like English or Dutch. These two categories were renamed by Pike (1945) as 'syllable-timed' and 'stress-timed' respectively. The differences in rhythm were attributed to the isochrony of the organizing unit in each category, i.e. syllables or interstress intervals (Abercrombie, 1967). A third category was added that described languages such as Japanese and Tamil, 'mora-timed' organized in morae (Ladefoged, 1975). In the 1980s, a different approach was proposed according to which the phonological features of the languages are what characterize their rhythm, for instance, syllabic complexity and vowel reduction (Dasher & Bolinger, 1982; Dauer, 1983). Languages were said to be organized on a continuum, where the two endpoints were syllable and stress-timed (Dauer, 1987). Ramus et al. (1999) measured in the speech signal vocalic and consonantal intervals and their durational variability to quantify the 'correlates of rhythm'. The hypothesis that listeners use differences in rhythm to discriminate languages has been extensively investigated in infants (e.g. Byers-Heinlein et al., 2010; Nazzi et al., 2000; Nazzi et al., 1998), adults (e.g. Ramus et al., 2003; Ramus & Mehler, 1999) and even non-human animals (Hauser et al., 2001; Toro et al., 2005). In contrast, intonation, a fundamental component of language prosody, has been somewhat neglected.

There have been numerous attempts to study and define intonation (Bolinger, 1982; Lieberman, 1958). The broad definition equates intonation to prosodic information, both rhythm and melody, whereas the narrow definition refers to the use of fundamental frequency for linguistic purposes (Allen, 1971; Levis, 1999). Pitch, as in the fundamental frequency of voiced sounds, is inextricably linked to loudness, timing patterns and often voice quality, complicating the distinction of prosodic features (Nolan, 2020; Roach, 2010). Here we refer to intonation as defined by Ladd, (2008, p.4) 'the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way'. Intonation has different functions: it is used to show attitude and the speaker's state (communicative function), to regulate discourse (turn-taking), to highlight different parts of an utterance/sentence (accentual function, prominence) and to signal phrase boundaries or type of sentences such as questions (grammatical function). Intonational systems differ across languages. At the word prosodic level, languages have been classified into three categories depending on whether words have tone, stress or lexical pitch accent (Beckman, 1986; Jun, 2010; Trubetzkoy, 1939). Languages like Chinese, Vietnamese and Thai use tone, English, Spanish and Catalan use stress accent, and Japanese, Basque and Ancient Greek use pitch accent. These different patterns are instantiated by different pitch movements. The marking of prominence in an utterance (postlexical) has been described in terms of falling and rising tones based on the direction of the pitch contour. These tones can be broken down into smaller constituents, the most prominent being the nucleus, which in languages like English coincides with the accented syllable of the most prominent word (highest fundamental frequency), usually in final position inside an intonational phrase. The nuclear accent can have different combinations of high and low tones across languages (Frota et al., 2007). Apart from prominence, intonation can also cue phrase boundaries with particular pitch patterns depending on the language (Jun, 2010). A phonological framework was created based on the work by Beckman & Pierrehumbert, (1986) and Pierrehumbert (1980) to annotate the different intonational patterns enabling comparisons across languages (Silverman et al., 1992). Therefore, different intonational patterns can be used to discriminate across languages.

There is mixed evidence concerning the effectiveness of intonational cues in adult language discrimination. Some of the languages used in the experiments are rather different in terms of rhythm and intonation such as English from Spanish, meaning that participants could be using either rhythm or intonational cues to discriminate the two (Atkinson, 1968). Ramus & Mehler (1999) resynthesized natural languages in a way that different types of phonological information were either kept or removed from the speech signal to test the contribution of different prosodic information. They tested if French participants (native speakers of a syllable-timed language) could discriminate English (stress-timed) from Japanese (mora-timed). Participants were able to distinguish the two languages when they had access to rhythm or both rhythm and intonation but failed when they only had access to intonation. Similarly, native speakers could identify above chance sentences in French (syllable-timed language) and English (stress-timed language) that were transformed into a laryngograph waveform that contains prosodic information (Maidment, 1976, 1983). Speakers of American English (stress-timed language), Cantonese Chinese (syllable-timed language) and Japanese (mora-timed language) speakers were able to identify sentences in the three languages when they were transformed into a pulse train signal that captures frequency, amplitude and timing patterns. The three languages not only differ in terms of their rhythm but also in their lexical accent, i.e. intonational use of stress, they are lexical stress, tone and pitch accent languages, respectively (Ohala & Gilbert, 1979). More recently, Komatsu et al. (2004) tested adults on their ability to identify Chinese, English, Japanese and Spanish. Spanish is a syllable-timed language that uses stress for lexical accent, similar to English. The sentences were edited into a pulse train that simulates only intonation. Participants were able to identify the languages above chance when the pair of the languages tested differed in their lexical accent, for instance Chinese from English, but failed when they use the same lexical accent, i.e. English and Spanish. When languages (or dialects) do not differ rhythmically, intonation appears to be enough of a cue in most cases. Supporting evidence stems from adults' studies testing native speakers of one of the languages tested. Participants had no trouble in discriminating British, American and Indian English (Bush, 1967) , Quebecois and European French (Menard et al., 1999) or Western from Eastern Arabic when they only had access to intonational cues due to low-pass filtering or pulse train (Barkat et al., 1999). The same was found for American English and German. In contrast, adults were unable to discriminate Australian from American English based only on intonation (Vicenik & Sundara, 2013). French adults (non-native) could not discriminate Catalan from Spanish when they had no access to intonational or segmental cues (Ramus et al., 2003). Thus, it could be that participants are always computing rhythmic and intonational information for each language, but they can use only one depending on what type of information is more useful.

Infants have also been tested on their abilities to discriminate languages using intonational cues. Ramus (2002) tested French newborns with Dutch and Japanese sentences that were resynthesized. The results showed that intonation was not enough of a cue for this between class discrimination. Previous research has shown that intonation is necessary for 7-month-old infants to discriminate between a pair of stress-timed languages, American English and German. Infants were able to discriminate when the stimuli were low-pass filtered, leaving behind prosodic information but such ability was impaired when intonation was replaced by a monotone (Chong et al., 2018). More recently, this ability was tested with a pair of syllable-timed languages, Western Catalan and Spanish, with even younger participants. Four-and-a-half-month-old infants were able to discriminate the languages when the sentences were natural speech, but when they were low-pass filtered discrimination was marginal. This led the authors to hypothesize that infants might need additional exposure to their native language to acquire the intonational patterns (Zacharaki & Sebastian-Galles, 2021).

The review of adult research points in the direction that the participant's native language may facilitate discrimination when only intonational cues are available (for a summary on the participants' identity and results on language discrimination and LID see tables 1, 2). There are two possibilities, first, participants are tested in their native language against a foreign one and second, both languages are foreign to participants. As mentioned, only native speakers can discriminate between two Arabic dialects (Barkat et al., 1999) or English and Japanese sentences (Ramus & Mehler, 1999). It is worth noting that in the second article, native participants were explicitly told that they would hear sentences in English versus an exotic language. Therefore, the role of familiarity and awareness that they would hear English were confounded, importantly nonnative participants were not given the same information. Converging results have also been found in studies investigating language discrimination in the visual modality. Spanish and/or Catalan speakers, but not English and Italian ones, were able to discriminate between silent clips of an actress reciting sentences in Spanish and Catalan (Soto-Faraco et al., 2007). Non-native English participants who had acquired this language before the age of 6 years could discriminate between silent clips of three speakers reciting English and French sentences. In contrast, participants who acquired English after 6 years could not distinguish between the languages (Weikum et al., 2013).

Here we test whether intonational patterns are enough to enable native and non-native speakers to discriminate between Western Catalan and Spanish. Previous research has shown that Spanish and

Catalan[13] belong to the same rhythmic class, leaving only intonational cues marking the differences between the two at the prosodic level (Prieto et al., 2012; Ramus et al., 1999; Ramus et al., 2003). Intonationally, Catalan and Spanish tend to behave similarly when compared with other Romance languages like Italian, but there are relevant differences between the two (Frota et al., 2007). Spanish is trochaic at the foot level, whereas Catalan is iambic (Gibson, 2010; Ohannesian, 2005; Oliva, 1992; Wheeler & Wheeler, 2005). In both languages the most frequent phrasing pattern is (S)(VO). However, (SV)(O) is also observed in Catalan, as a means of maintaining the constituents of equal duration. This pattern is rare in Spanish (Imperio et al., 2009). Finally, Catalan and Spanish mark boundaries using different cues, for instance pitch reset is used more frequently in Spanish than in Catalan, while the opposite holds true for preboundary lengthening (Frota et al., 2007). We will test participants with low-pass filtered sentences in Western Catalan and Spanish. Low-pass filtering eliminates most segmental information, though some vowel information might be left in the signal, and leaves prosodic features mostly intact. Eastern Catalan (the one previously used in the majority of experimental studies and spoken in the Barcelona area) differs in a significant way from Spanish in the distribution of vowels. For instance, mid-vowels are very frequent in Spanish (49%) but not so frequent in Eastern Catalan (16%). However, Western Catalan, like Spanish, has no vowel reduction (Carbonell & Llisterri, 1992; Wheeler & Wheeler, 2005), which results in them having a comparable vowel distribution (for a full description see Zacharaki & Sebastian-Galles, 2021). Therefore, the fact that some vocalic information is left with low-pass filtered stimuli could not be used as a cue to discriminate between these two languages. We tested two groups of adult participants, one of native speakers (Catalan and/or Spanish speakers) and another one of non-native speakers (natives of several different languages). Based on previous research, our hypotheses are that intonation will be enough to discriminate between the two languages and that native speakers will outperform non-native speakers.

[13] Eastern (Central) and Western (North-western) Catalan dialects are similar intonationally which is why we refer to them as Catalan in the description of their intonational features (Prieto et al., 2015).

<table>
<tr><th colspan="7">Table 1. Summary of previous results on Language discrimination with adults</th></tr>
<tr><th>Experiment</th><th>Language pair</th><th>Manipulation</th><th>Segment diffs</th><th>Rhythm diffs</th><th>Participants' L1</th><th>Discrimination</th></tr>
<tr><td rowspan="4">Ramus, Dupoux, & Mehler (2003)</td><td>English, Spanish</td><td rowspan="4">Resynthesis only rhythm (flat sasasa)</td><td>No</td><td>Yes</td><td rowspan="4">French</td><td>Yes</td></tr>
<tr><td>Catalan, English</td><td>No</td><td>Yes</td><td>Yes</td></tr>
<tr><td>English, Dutch</td><td>No</td><td>No</td><td>No</td></tr>
<tr><td>Catalan, Spanish</td><td>No</td><td>No</td><td>No</td></tr>
<tr><td rowspan="2">Vicenik, & Sundara (2013)</td><td rowspan="2">American English, German</td><td>Low-pass filtering</td><td>No/Limited</td><td>No</td><td rowspan="2">American English</td><td>Yes</td></tr>
<tr><td>Resynthesis- Only rhythm</td><td>No</td><td>No</td><td>Yes</td></tr>
</table>

		Resynthesis- Only intonation	No	No		Yes
	American English, Australian English	Low-pass filtering	No/Limited	No	American English	Yes
		Resynthesis- Only rhythm	No	No		Yes
		Resynthesis- Only intonation	No	No		No

Table 2. Summary of previous results on Language identification with adults

Experiment	Languages	Manipulation	Segment diffs	Rhythm diffs	Participants’ L1	Identification
Atkinson (1968)	English,	Low-pass	No/Limited	Yes	-	Yes

	Spanish	filtering				
Barkat, Ohala, & Pellegrino (1999)	Eastern Arabic, Western Arabic	Natural	Yes	No	Western Arabic	Yes
					Non-native speakers	Yes
		Sinusoidal pulses	No	No	Western Arabic	Yes
					Non-native speakers	No
Bush (1967)	American, British, Indian English	Low-pass filtering	No/Limited	No	American English	Yes
Komatsu, Arai, & Sugawara (2004)	Chinese, English	Pulse train simulating f0	No	No	-	Yes
	Chinese, Japanese					Yes
	Chinese, Spanish					Yes

	English, Japanese					Yes
	English, Spanish					No
	Japanese, Spanish					Yes
Maidment (1983)	French, English	Laryngograph waveform	No	Yes	English	Yes
Menard, Ouellon, & Dolbec (1999)	European French, Quebecois French	Low-pass filtering	No/Limited	No	Quebecois French	Yes
Moftah, & Roach (1988)	Arabic, English	Laryngograph waveform	No	No	Arabic or English	Yes
		Low-pass filtering	No/Limited	No	Arabic or English	Yes

Muthusamy, Barnard & Cole (1994)	English, Farsi, French, German, Japanese, Korean, Mandarin Chinese, Spanish, Tamil, Vietnamese	Natural	Yes	Yes	10 English speakers and 2 from each Language other than English	Yes
Ohala & Gilbert (1979)	English, Cantonese Chinese, Japanese	Triangular pulse train	No	Yes	English, Chinese, Japanese	Yes
Ramus & Mehler (1999)	English, Japanese	Resynthesis (‘aaaa’)	No	No	French	No
					American English	Yes

3.3. Methods

3.3.1. Participants

One hundred and twenty-eight participants were tested online. Participants were organized into two groups based on their native language, i.e. the language they were exposed to during the first year of life. The native group consisted of 64 participants, who either heard Spanish and/or Catalan from birth (M_{age}=29, SD=9, range= 19-65, 41 females, 20 male, 2 non-binary and 1 preferred not to answer). The vast majority of the native speakers of Catalan spoke the Eastern dialect (the dialect of the Barcelona area); but as Eastern and Western Catalan are not different in terms of intonation (Prieto et al., 2015) we considered speakers of different dialects of Catalan as equivalent for the purpose of this experiment.

The non-native group consisted of 64 participants, who did not have any exposure to Catalan or Spanish during the first 5 years of life [14] (M_{age}=31, SD=7, range= 21-65, 38 females, 25 male, 1 preferred not to answer). The participants in the non-native group came from different linguistic backgrounds, mostly European languages (for more information see Appendix A.). To ensure that previous knowledge of Catalan and or Spanish did not affect the responses of the participants, we asked them to self-evaluate their comprehension in each language. Out of the 64 non-native participants, 12 reported having no comprehension of the two languages and the other 52 had on average 1.2 for Catalan and 2.6 for Spanish on a scale of 5 maximum per language.

[14] Two participants were exposed to Spanish before 5 years of age but had no exposure during the first year of life which is why we included them in the non-native group. Exposure was very limited, and comprehension was low. The pattern of results reported in the results section does not change if these two participants are removed or reclassified as natives.

Participants declared having no hearing or learning disabilities and only participants who completed the whole experiment were included in the final sample. An additional group of 7 participants participated in the experiment but were not included in the final sample for the following reasons: use of a mobile phone (n=2), underaged (n=2), failed to complete the test (n=2), technical issue (n=1). Participants were recruited through the database of adult participants from the ‘Laboratoris de Neurociència’ at the Universitat Pompeu Fabra and through the internet. The experiment was promoted on social media, i.e. Facebook, in groups of expatriates and through advertisements and distribution lists in different Spanish and foreign Universities (in particular the University of Athens). Participants’ recruitment informed potential participants that the instructions of the experiment would be in English, so minimal knowledge of this language was required. Participation was completely voluntary, and participants did not receive any monetary compensation. The experiment reported in this article was conducted in accordance with the principles expressed in the Declaration of Helsinki and approved by the local ethical committee, the clinical research ethical committee of the Parc de la Salut Mar.

3.3.2. Stimuli

We used the sentences in Western Catalan and Spanish from experiment 4 in Zacharaki & Sebastian-Galles (2021). A highly proficient bilingual speaker with no noticeable accent in either language recorded 12 sentences in each language. The stimuli were recorded using an Audio-Tecnica microphone (AT2050) at a sampling rate of 44,100 Hz. The recording took place in a soundproof room at the ‘Laboratori de Recerca en Infancia’ at the Universitat Pompeu Fabra using Audacity (Mazzoni, 1999). The sentences were inspired by the children’s storybook “Frog, where are you?” (Mayer, 1969). The sentences were produced in an infant-directed manner at a natural speaking rate. The speaker was instructed to avoid unnecessary pauses. Sentences were analyzed using Praat and sentences did not differ in terms of number of syllables and pitch (table 3) (Boersma & Weenink, 2019). We low-pass filtered the sentences at a threshold of 400Hz with a roll-off of 48dB/octave using Audacity, in this way, we eliminated most of the segmental information leaving prosody intact.

Table 3. Acoustic properties (mean and standard deviations in parentheses) of the low-pass filtered sentences

	Western Catalan	**Spanish**
Average number of syllables	17 (2.7)	18 (2.3)
Average duration of sentences	3.8 sec (0.6)	3.8 sec (0.3)
Mean f0 of sentences	220 Hz (14)	215 Hz (12)
Mean f0 range	181 Hz (36)	210 Hz (52)

3.3.3. Procedure

We adapted the procedure used in Ramus & Mehler (1999) to an online set-up using JavaScript. Participants accessed the experiment from a direct URL link (https://cbclab-online.upf.edu/sap/kzacharaki/discri_lang_online/index.php?id=&lang=en). In the advertisements of the experiment, it was made clear that participants would need a computer and access to the internet to participate. Participants were informed that their responses would be registered and that their participation would not be compensated. Participants read the consent form by clicking on the screen and provided their consent by clicking that they had read the form and that they gave their consent to use the data for our investigation. The instructions described that a group of ethnolinguists went to the Amazonian jungle to investigate two exotic languages called 'Zahatu'[15] and 'Moltec'[16]. The researchers recorded indigenous

[15] We used the label Zahatu instead of Sahatu as in the original paper by Ramus & Mehler (1999) , because Z and M are symmetrically placed on the QWERTY keyboard.

people saying representative sentences in the two languages. Since they had no access to the internet, they only had hard copies of the recordings, which were degraded in the flight back. Participants were informed that their task was to classify each sentence as Zahatu or Moltec using the **Z** and **M** keys from their keyboards, respectively. Participants heard a low-pass filtered sentence in Greek, which was a direct translation from one of the sentences used in the experiment, to give them a chance to adjust the volume of their device to a comfortable level and to get familiarized with the properties of the experimental stimuli ("short and of poor quality"). They were informed that they would listen to one sentence of each language as familiarization, followed by the rest of the sentences. Participants were told that they would get feedback about their performance after each response. They were asked to classify every sentence even if they were unsure of their response. They were also asked to stay focused and avoid noisy environments.

The experiment was organized into two phases: training and test (although from the point of view of the participant there was no differentiation between training and test phases). Twenty sentences, 10 in Catalan and 10 in Spanish were presented in each part of the experiment. The sentences used in each phase were different. Since we had less sentences recorded than in the original article from Ramus & Mehler (1999), we decided to repeat some of the sentences. In the training phase we used 5 sentences in each language repeated twice, yielding two blocks. In this manner we ensured that none of the sentences was repeated consecutively. The order of the sentences in each block and language was randomized with the only restriction that the last sentence of the first block and the first sentence of the second block were not the same and no more than 3 sentences in the same language would be presented in a

[16] The labels Zahatu and Moltec were randomized across participants, in other words some participants heard the Catalan sentences which were labelled as Zahatu and for other participants it was labelled as Moltec and vice versa for the Spanish ones.

row. The sentence used as familiarization before the training phase was always presented last in the first block to minimize memory effects.

If the participants' responses reached 70% in the training phase, the test phase started. Otherwise, the training was repeated up to 3 times or until they reached the 70% success criterion. If participants failed to reach the 70% threshold after the 3 repetitions, the test phase started. The test phase had the same structure as the training one with the only difference that participants were presented 7 sentences in each language and 3 out of the 7 were repeated, yielding a total of 10 sentences per language. The same restrictions on the randomization of the sentences applied as in the training phase. Participants were given feedback on their answers in both cases with a check symbol (✓) or an X.

At the end of the experiment, participants were asked to describe how they discriminated between the two languages, and they were redirected to a Google form by clicking on a link, where they had to fill in their demographic information and some details on their language background. They were asked to indicate the age of acquisition from a list of languages. The options were: Spanish, Catalan, Galician, Basque, English, French, Italian, German, Other1, Other2. If they chose Other, they had to specify in a blank space which language they referred to. The age options we offered were in years: first year of life, 2-5, 5-10, 10-15, 15-20, 20-25, after 25. Participants were given a 'don't know' option if they did not know the age of acquisition. Lastly, they were asked to evaluate their comprehension from 0 to 5 for the same list of languages as before: 0 corresponding to unknown and 5 to native. The whole session lasted about 15 minutes. For each participant 3 files were saved at the safe repository of the Universitat Pompeu Fabra. One excel file contained information about the stimuli presentation, a text file contained technical information on the device used for the experiment and the third file contained the responses from the language questionnaire.

3.4. Results

The raw data are summarized in table 4 showing the percentage of correct responses separately for each part of the experiment, i.e. training(s) and test. Non-native participants performed more training blocks than native ones, the former completed 142 blocks of training in total while the latter completed 133. Statistical analyses were performed on the responses from the test session. Responses to 'Zahatu' that were correctly identified were considered as hits and incorrect responses as misses. Responses to 'Moltec' that were correctly rejected, were considered as correct rejections and incorrect responses as false alarms. A' scores were obtained in Rstudio (RStudio Team, 2019) using a script by Pallier (2002). A' scores are a non-parametric equivalent of d' scores and they range from 0 to 1; chance level discrimination is at 0.5. Figure 3.1. shows the individual A' scores plotted separately for the two groups of participants. We ran two separate t-tests for each group of participants to see if their A' scores were higher than chance levels. Natives ($t(63) = 7.96$, $p < .001$, $d = 1.00$) and non-natives ($t(63) = 4.00$, $p < .001$, $d = 0.50$) had A' scores above chance. A one-way ANOVA with nativeness (native, non-native) as a between participants factor revealed significant differences between the two language groups ($F(1, 126) = 6.05$, $p = .015$, $\eta^2 p = .05$). On average native participants responded more accurately than non-natives ($M_{nat}=0.71$, $M_{non}=0.61$).

Table 4. Numbers represent the participants that completed each training phase and moved directly to the test. Percentage of correct responses are shown in parentheses.

Participants	1st training	2nd training	3rd training	test
Native	23 (67%)	13 (66%)	28 (64%)	64 (66%)
Non-native	20 (62%)	10 (60%)	34 (60%)	64 (59%)

We also calculated the equivalent, non-parametric response bias B''_D (Donaldson, 1996). Zero stands for no bias, positive values for a conservative bias and negative ones for a liberal bias. The bias measurement, B''_D showed that the two groups had no significant differences in their response bias ($t(124)$[17] $= -0.78$, $p = .435$, $d = -0.14$). Response biases from both groups of participants are very close to zero ($M_{nat}=-0.0368$, $M_{non}=0.0365$).

We analysed the self-reported comprehension of Catalan and Spanish. We tested the variance in the two groups, native (Catalan and/or Spanish), and non-native, which was significantly different for the comprehension scores of Spanish ($p < .001$) but not for Catalan ($p=0.38$). An independent two-sample t-test of unequal variance indicated that Spanish comprehension scores differ significantly ($t(63.78) = -11.99$, $p < .001$, $d = -2.12$) across the two groups (natives: M=4.9, SD=0.1, non-natives: M=2.6, SD=1.5). An independent two sample t-test of equal variance indicated that Catalan comprehension scores also differ significantly ($t(126) = -10.36$, $p < .001$, $d = -1.83$) across the two groups (natives: M=4, SD=1.6, non-natives: M=1.2, SD=1.4). We computed two spearman rank correlations (rho) to assess if there is a relationship between the self-reported comprehension of Spanish or Catalan and the A' scores of the non-native participants. There was no significant correlation between comprehension scores of Spanish and A' scores ($r(62)=0.16$, $p=0.19$), nor between comprehension scores of Catalan and A' scores ($r(62)=0.20$, $p=0.10$) (see figure 3.2. A and B respectively).

Out of the 128 participants, 62 (natives 42, non-natives 20) responded to the debriefing question on how they classified the sentences. The most common answers in the native group were similarity to Catalan or Spanish (n=20), marked use of /r/ (n=12) and intonation (n=8). Similarly, the most common answers in the non/native group were the marked use of /r/ (n=9), similarity to

[17] Two of the 128 participants had a perfect Hit Rate score (Hits/(Hits+Misses)), equal to 1, which resulted in no response bias (NaNs) and were excluded for the comparison. One was a native speaker, and one was a non-native one.

Catalan or Spanish (n=5) and intonation (n=4). The remaining two participants in each group provided other anecdotal explanations.

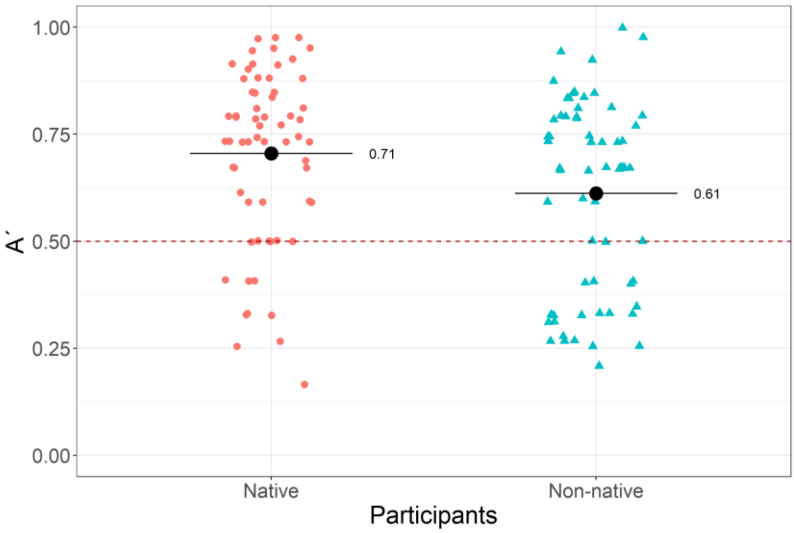
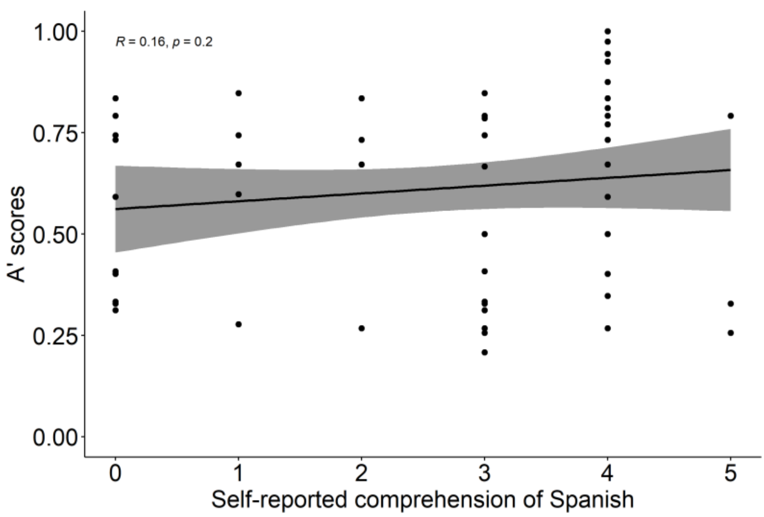


Figure 3.1. Individual A' scores are plotted in dots for the Native and in triangles for the Non-native participants. The black dots with the vertical lines stand for the mean in each group. The dotted red line represents chance level at 0.5.

A.



B.

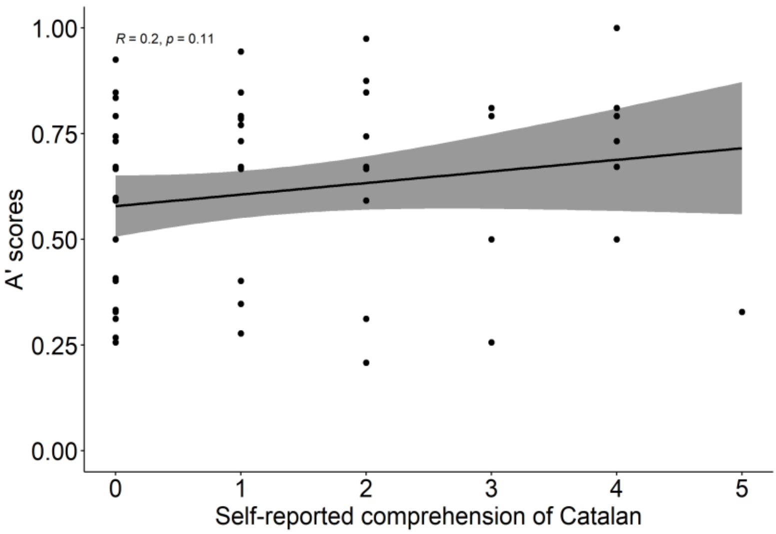


Figure 3.2. Self-reported comprehension scores for the non-native participants in Spanish (A.) and Catalan (B.) are plotted in relation to A' scores obtained during the test.

3.5. Discussion

In this study, we found that adults can discriminate between sentences in Western Catalan and Spanish using only intonational cues. The analysis of the A' scores showed that both groups of participants could discriminate above chance. It also uncovered statistically significant differences between the two groups: natives performed better than non-natives. Self-reported comprehension of Spanish or Catalan in the non-native group did not correlate with the discrimination score.

The contribution of prosodic information to language identification has also attracted attention in the domain of automatic language identification. Although much research has been conducted since the 1960's on automatic LID (Komatsu, 2007; Singh & Poonia, 2015), humans are the best language identification system currently, given that they can respond within a second in a language identification task (Muthusamy et al., 1994). Having access to both segmental and prosodic information yields the highest identification scores, but it has been found that prosodic information alone is enough to discriminate typologically distinct languages in automatic systems (Cummins et al., 1999; Rouas, 2007; Thymé-Gobbel & Hutchins, 1996).

As aforementioned, the role of intonation in language discrimination has yielded conflicting results. When the languages compared are rhythmically dissimilar there is evidence that intonation alone is not enough to enable discrimination (Ramus & Mehler, 1999), while others suggest the contrary (Komatsu et al., 2004). Other articles also provide evidence that intonation is important but they did not disentangle it from rhythm (Atkinson, 1968; Maidment, 1976, 1983; Ohala & Gilbert, 1979). When the languages compared are rhythmically similar, intonation has been shown to be enough of a cue in most cases (Barkat et al., 1999; Bush, 1967; Menard et al., 1999). Vicenik & Sundara (2013) reported contradictory results using re-synthesized speech where all segmental information is replaced by a pitch-modulated stream of 'aaaa'. American English participants could discriminate American English from German above chance but not their native dialect from Australian English. It is worth noticing that in this experiment participants' responses were close to chance levels. It may be argued that this type of experimental manipulation renders the stimuli as non-speech as it sounds more like somebody humming than actually speaking. Our findings fit in with the majority of previous research on rhythmically similar discrimination, showing that intonational cues are enough for adults to discriminate between Western Catalan from Spanish.

Discrimination scores obtained by our participants, in general, were relatively low, even in the native group, highlighting that it is a difficult task and suggesting that adults are very keen on relying on segmental information. Previous research corroborates the importance of segmental information. Adults were able to discriminate between Swiss German dialects using segmental information, when rhythm and/or intonation were swapped across the dialects tested (Leemann et al., 2018). Also, our decision to use the labels ‘Zahatu’ and ‘Moltec’ in order to equalize the task among natives and non-natives made the task harder. When participants are provided with a label that corresponds to their native language, they perform better (Ramus & Mehler, 1999).

Infants at 4.5 months of age can discriminate between Western Catalan and Spanish with natural sentences. However, when the sentences were low-pass filtered, same stimuli we used here, we found only a marginal effect of discrimination ($p = .066$, Experiment 4). Infants of the same age can easily discriminate between languages or dialects that differ at the segmental level or in terms of rhythm. This marginal discrimination based on intonational cues pointed towards the possibility that infants need more exposure to their native language in order to acquire its intonational patterns (Zacharaki & Sebastian-Galles, 2021). Supporting evidence comes from Chong et al. (2018) who showed that intonation is critical for language discrimination at 7 months. Despite adults relying heavily on segmental information to perform language discrimination, they seem to be better than young infants. This is somehow striking as prosodic information (rhythmic and intonational) is considered the primary source of information in language processing in the very first months of life (Abboub et al., 2016; Gervain, 2018). Though, the number of studies is very low and the matter deserves further investigation.

Evidence on the role of previous experience with the languages of test in studies of language discrimination is very limited. Barkat et al. (1999) tested the discrimination of two dialects of Arabic (Western and Eastern) by native Arabic participants and by non-native ones. Their results also showed that native participants outperformed non-native ones. It is worth noticing that most non-native participants in our investigation were able to understand some Spanish and/or Catalan, also that we found no correlation between comprehension of Spanish or Catalan and discrimination. Although our results have to be considered with a lot of caution, as we had very little control on participants' selection, they are compatible with the hypothesis that early exposure (corresponding here to nativeness) entails some kind of fine-grained knowledge of intonational properties that allows native participants to better perform the task. Our results do not allow us to conclude if native and non-native participants used different types of information to perform the task. Native and non-native participants provided similar explanations to describe how they performed the task. However, how reliable such explanations are is debatable. For instance, many native and non-native speakers indicated that the sound /r/ sounded more marked in one language than in the other. The realization of this phoneme is exactly the same in the two languages (and we checked that they sounded equivalent in our recordings), low-pass filtering significantly distorts it, and as said, performance was not very accurate in general. Our research is consistent with results showing early exposure enhances visual language discrimination (of talking faces without any sound) (Soto-Faraco et al., 2007; Weikum et al., 2013). Soto-Faraco et al. (2007) reported that only Spanish and Catalan speakers, but not Italian or English speakers, could visually discriminate Catalan from Spanish. Weikum et al. (2013) reported that visual discrimination of French and English was only possible for non-native speakers if English had been acquired before the age of 6.

Before concluding, it is worth commenting on the use of low-pass filtering as an effective tool to remove segmental information. Although, it was been used extensively in the field of language discrimination both with adults and infants, it is not without relevant limitations. By cutting-off frequencies above 400Hz, pitch modulations above the threshold may be eliminated, while at the same time some segmental information presented in low frequencies may be preserved . Eliminating pitch modulations is more likely to happen for female speakers with high fundamental frequencies. Ramus & Mehler (1999) mention that by listening to low-pass filtered speech, segmental information can be discerned, although it is highly degraded. As already said, such remarks coincide with some of the answers collected in the debriefing question in our experiment. Despite the criticisms and limitations of low-pass filtering, the technique has uncovered differences between natives and non-natives. In the future, we could run a control experiment with stimuli that have no intonational cues. There two different ways of doing so: either removing the pitch contours from the low-pass filtered sentences and applying a monotone one or resynthesizing the sentences to remove segmental information and superimpose the intonational contours of a third language (Ramus, 2002; Seidl, 2007). If we no longer observe discrimination, we could conclude that intonation is necessary for adults to discriminate Western Catalan from Spanish.

To sum up, we found that adults are able to discriminate and classify sentences in Western Catalan and Spanish using intonation. Native speakers outperformed the non-natives speakers, but both groups of participants were above chance level. Our results extend previous findings by testing a new pair of languages that has not been used in the past with adults. Future research needs to replicate our findings. Our findings can provide valuable information for automatic language identification on the use of intonation.

3.6. References

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, *162*, 46–59. https://doi.org/10.1016/j.bandl.2016.08.002

Abercrombie, D. (1967). *Elements of general phonetics*. University Press. https://books.goog le.es/books?id=bmQhAQAAMAAJ

Allen, V. F. (1971). Intonation from Theory to Practice Teaching. *TESOL Quarterly*, *5*(1), 73–81.

Atkinson, K. (1968). Language Identification from Nonsegmental

Cues. *The Journal of the Acoustical Society of America*, *44*(1), 378–378. https://doi.org/10.1121/1.1970573

Barkat, M., Ohala, J., & Pellegrino, F. (1999). Prosody as a distinctive feature for the discrimination of arabic dialects. *6th European Conference on Speech Communication and Technology (EUROSPEECH'99)*.

Beckman, M. E. (1986). *Stress and non-stress accent*. De Gruyter Mouton.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255–309. https://doi.org/DOI: 10.1017/S095267570000066X

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. http://www.praat.org/

Bolinger, D. (1982). Intonation and Its Parts. *Language*, *58*(3), 505–533.

Bush, C. N. (1967). Some Acoustic Parameters of Speech and Their Relationships to the Perception of Dialect Differences. *TESOL Quarterly*, *1*(3), 20. https://doi.org/10.2307/3586196

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343–348. https://doi.org/10.1177/0956797609360758

Carbonell, J. F., & Llisterri, J. (1992). Catalan. *Journal of the International Phonetic Association*, *22*((1-2)), 53–56. https://doi.org/10.1017/S0025100300004618

Chong, A. J., Vicenik, C., & Sundara, M. (2018). Intonation Plays a Role in Language Discrimination by Infants. *Infancy*, *23*(6), 795–819. https://doi.org/10.1111/infa.12257

Cummins, F., Gers, F., & Schmidhuber, J. (1999). *Automatic discrimination among languages based on prosody alone*. *IDSIA-03-99*.

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, *12*(2), 58–71. https://doi.org/doi:10.1017/S0025100300002462

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed.

In *Journal of Phonetics* (Vol. 11, Issue 1, pp. 51–62). Elsevier Science.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 447–450).

Donaldson, W. (1992). *Measuring Recognition Memory*. *121*(3), 275–277.

Frota, Sonia, Imperio, M., Elordieta, G., Prieto, P., Frota, S., Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. (2007). The phonetics and phonology of intonational phrasing in Romance. In *Segmental and prosodic issues in Romance Phonology* (pp. 131–154). John Benjamins. hal-00142940

Gervain, J. (2018). Gateway to language: The perception of prosody at birth. In *Boundaries crossed, at the interfaces of morphosyntax, phonology, pragmatics and semantics* (pp. 373–384). Springer.

Gibson, M. (2010). *Fonología del español: enfoque desde la teoría de optimidad. Issues in Spanish Phonology from an Optimality-Theoretic Approach*. Universidad Complutense de Madrid.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-64. https://doi.org/10.1016/s0010-0277(00)00132-3

Imperio, M. D., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2009). *Intonational phrasing in Romance: The role of syntactic and prosodic structure*. *Selkirk 2000*. https://doi.org/10.1515/9783110197587.1.59

Johnson, E. K., Jusczyk, P. W., & Ramus, F. (2003). The role of segmental information in language discrimination by English-learning 5-month-olds. *Jusczyk Lab Final Report*. http://hincapie.psych.purdue.edu/Jusczyk

Jun, S.-A. (2010). Prosodic Typology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press: Oxford.

https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Komatsu, M. (2007). Reviewing human language identification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4441 LNAI*, 208–228. https://doi.org/10.1007/978-3-540-74122-0_17

Komatsu, M., Arai, T., & Sugawara, T. (2004). *Perceptual Discrimination of Prosodic Types and Their Preliminary Acoustic Analysis*.

Komatsu, M., Mori, K., Arai, T., Aoyagi, M., & Murahara, Y. (2002). Human language identification with reduced segmental information. *Acoustical Science and Technology*, *23*(3), 143–153. https://doi.org/10.1250/ast.23.143

Ladd, D. R. (2008). *Intonational_Phonology.pdf* (C. U. Press (ed.); June 2012). https://doi.org/https://doi.org/10.1017/CBO9780511808814

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich : New York. https://books.google.es/books?id=w_5YAAAAMAAJ

Leemann, A., Kolly, M. J., Nolan, F., & Li, Y. (2018). The role of segments and prosody in the identification of a speaker's dialect. *Journal of Phonetics*, *68*, 69–84. https://doi.org/10.1016/j.wocn.2018.02.001

Levis, J. M. (1999). Intonation in Theory and Practice, Revisited. *TESOL Quarterly*, *33*(1), 37. https://doi.org/10.2307/3588190

Lieberman, P. (1958). *Intonation, perception, and language*. Massachusetts Institute of Technology.

Lloyd James, A. (1940). *Speech signals in telephony,*. Sir I. Pitman & Sons.

Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and Hearing: Work in Progress*, *2*(74).

Maidment, J. A. (1983). Language recognition and prosody: further evidence. *Speech, Hearing and Language: Work in Progress*, *1*, 133–141.

Mayer, M. (1969). *Frog, where are you?* Dial Books for Young Readers.

Mazzoni, D. (1999). *Audacity®*.

Menard, L., Ouellon, C., & Dolbec, J. (1999). Prosodic markers of regional group membership: the case of the French of Quebec versus France. In J. Ohala (Ed.), *proceedings of the xivth international congress of phonetic sciences (icphs 99')* (pp. 1601–1604).

Moftah, A., & Roach, P. (1988). Language recognition from distorted speech : Comparison of techniques. *Journal of the International Phonetic Association*, *18*(1988), 50–52. https://doi.org/10.1017/S0025100300003558

Muthusamy, Y. K., Barnard, E., & Cole, R. A. (1994). Reviewing Automatic Language Identification. *IEEE Signal Processing Magazine*, *11*(4), 33–41. https://doi.org/10.1109/79.317925

Nazzi, T, Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language*, *43*, 1–19.

Nazzi, Thierry, Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 24, Issue 3, pp. 756–766). American Psychological Association. https://doi.org/10.1037/0096-1523.24.3.756

Nolan, F. (2020). Intonation. In B. Arts, A. McMahon, & L. Hinrichs (Eds.), *The Handbook of English Linguistics, Second Edition* (Second Edi, pp. 385–405).

Ohala, J. J., & Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody. *Problèmes de Prosodie: Expérimentations, Modèles et Fonctions*, *2*, 123.

Ohannesian, M. (2005). *La asignación del acento en castellano*. Universitat Autònoma de Barcelona,.

Oliva, S. (1992). *La mètrica i el ritme de la prosa*. Quaderns Crema.

Pallier, C. (2002). *Computing discriminability and bias with the R software*. 1–6.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. *RELC Journal*, *15*(1), 117–118. https://doi.org/10.1177/003368828401500113

Pike, K. L. (1945). *The Intonation of American English*. http://search.ebscohost.com/login.aspx?direct=true&db=&AN=&site=eds-live

Prieto, P, Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M. D. M. (2015). Intonational phonology of Catalan and its dialectal varieties. In *Intonation in Romance* (pp. 9–62).

Prieto, P, Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*(6), 681–702. https://doi.org/10.1016/j.specom.2011.12.001

Ramus, F, Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *15th ICPhs Barcelona*, *May 2014*, 1–6. http://cogprints.org/3079/

Ramus, F, & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512–521. http://www.ncbi.nlm.nih.gov/pubmed/9921674

Ramus, F, Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Ramus, F (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. In *Annu. Rev. Lang. Acquis* (Vol. 2). https://doi.org/10.1075/arla.2.05ram

Roach, P. (2010). *English Phonetics and Phonology Fourth Edition: A Practical Course*. Cambridge University Press. https://books.google.es/books?id=2LPl3_8xM6kC

Rouas, J. L. (2007). Automatic prosodic variations modeling for language and dialect discrimination. *IEEE Transactions on Audio, Speech and Language Processing*, *15*(6), 1904–1911. https://doi.org/10.1109/TASL.2007.900094

RStudio Team. (2019). *RStudio: Integrated Development Environment for R RStudio, Inc.* http://www.rstudio.com/

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48. https://doi.org/10.1016/j.jml.2006.10.004

Silverman, K. E. A., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P. J., Pierrehumbert, J. B., & Hirschberg, J. (1992). TOBI: a standard for labeling English prosody. *ICSLP*.

Singh, N. K., & Poonia, A. S. (2015). Analysis of Prosody based Automatic LID Systems. *International Conference on Communication, Control and Intelligent Systems (CCIS)*, 2–6.

Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception and Psychophysics*, *69*(2), 218–231. https://doi.org/10.3758/BF03193744

Thymé-Gobbel, A. E., & Hutchins, S. E. (1996). On using prosodic cues in automatic language identification. *ICSLP*, 1768.

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2005). Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 95–100. https://doi.org/10.1037/0097-7403.31.1.95

Trubetzkoy, N. S. (1939). *Grundzüge der phonologie*. De Gruyter.

Vicenik, C., & Sundara, M. (2013). The role of intonation in language and dialect discrimination by adults. In *Journal of Phonetics* (Vol. 41). https://doi.org/10.1016/j.wocn.2013.03.003

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience*, *7*(November), 86. https://doi.org/10.3389/fnsys.2013.00086

Wheeler, M. W., & Wheeler, R. L. M. W. (2005). *The Phonology of Catalan*. OUP Oxford. https://books.google.gr/books?id=PkoVDAAAQBAJ

Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. *Cognition*, *February*, 104628. https://doi.org/10.1016/j.cognition.2021.104628

3.7. Appendix A.

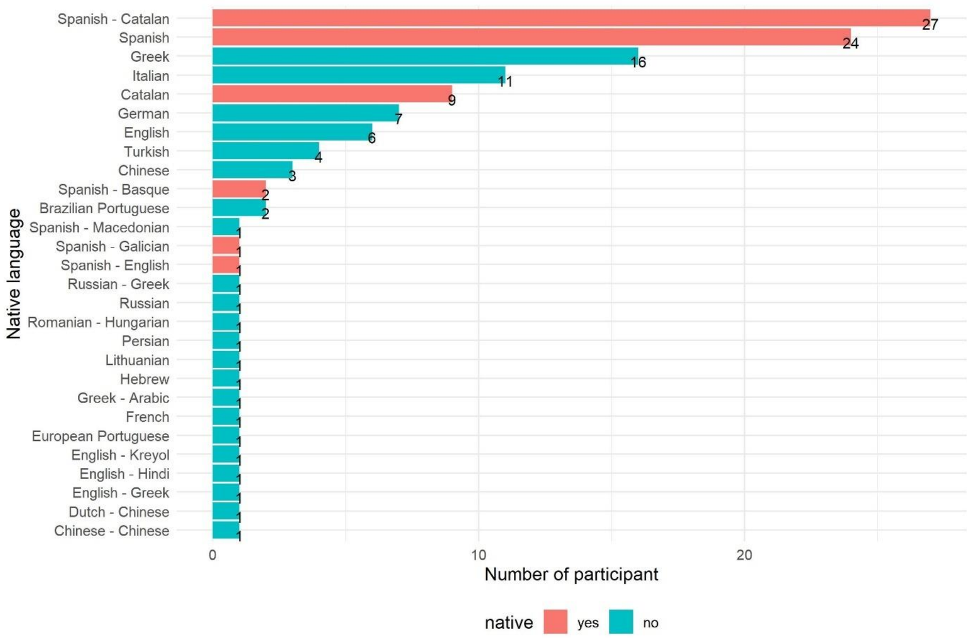


Figure 3.1. Details on the language profile of the participants are depicted in the bar plot. Participants are classified according to the language(s) they heard during their first year of life. Bars in coral represent the native group and bars in blue the non-native

one.

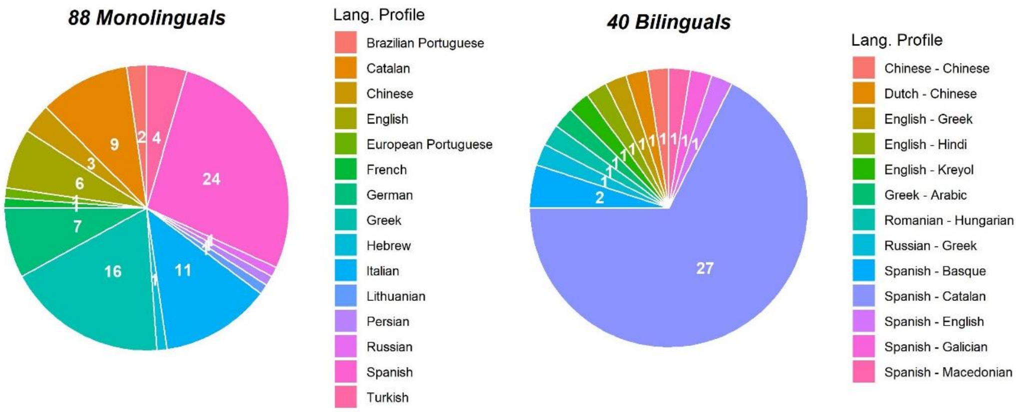


Figure 3.2. Participants are plotted as monolinguals or bilinguals depending on the language(s) heard during the first year of life.

4. Before perceptual narrowing: The emergence of phonetic representations

Konstantina Zacharaki[*], Nuria Sebastian-Galles[*]

[*] Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Barcelona, Spain

4.1. Abstract

The present study investigates the precursors of representations of phonemes in 4.5-month-olds. The emergence of phonemes has been mainly studied within the framework of perceptual narrowing, i.e. infants tuning to their native language and losing sensitivity to non-native speech. One of the mechanisms behind this phenomenon is distributional learning. In this article, we tested the preference of 4.5-month-old infants using lists of pseudowords that resemble the vowel distribution of the native or a non-native language. We found that infants prefer listening to the lists mirroring the native language. The results suggest that infants can extract vowel information from novel stimuli, and they can map it on pre-existing knowledge on vowels that leads to a preference for the native lists.

Keywords: phonemes, vowels, distributional learning, proto-segmental information, infants, language perception

4.2. Introduction

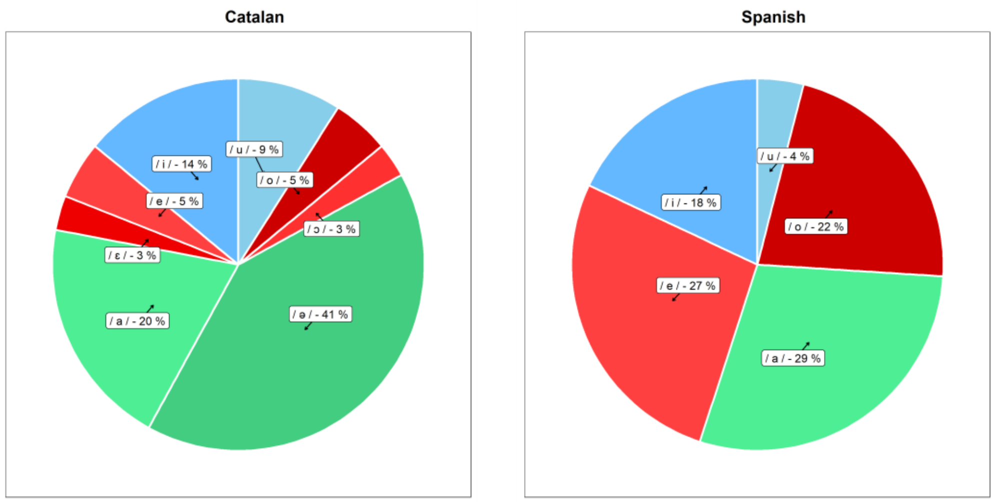
One fundamental milestone in early language acquisition is the establishment of the phoneme categories of the language of exposure. Initially newborns are able to discriminate virtually all phoneme contrasts existing in the different languages of the world, even if never exposed to them or their caregivers cannot produce them (Cheour-Luhtanen et al., 1995; Eimas et al., 1971; Trehub, 1976; for reviews see Aslin et al., 1998; Werker & Gervain, 2013). This initial capacity starts to decline around 6 months of age, when infants' perception of the native categories improves and perception of the non-native ones declines: the so-called perceptual narrowing. Perceptual narrowing has been extensively reported, taking place first for lexical stress and tone (Yeung et al., 2013), then for vowels (Kuhl et al., 1992; Polka & Werker, 1994; for a meta-analysis see Tsuji & Cristia, 2013) and later for consonants (Werker & Tees, 1984). Perceptual narrowing of vowels has been reported for Spanish-learning infants between 4.5 and 8 months of age. Bosch & Sebastián-Gallés (2003) showed that 4.5 Spanish-learning infants can discriminate the Catalan-specific /e/ – /ɛ/ vowel contrast, that Spanish native adults, as well as 8-month-old Spanish learning infants no longer discriminate (Bosch et al., 2000; Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999 among others). In contrast, Catalan natives (at 4.5 and 8-months of age) have no difficulties to perceive this contrast.

There is a generalized consensus in that infants' sensitivity to the distributional properties of the speech sounds is critical in the perceptual narrowing. The emergence of phonetic categories would result from infants' computation of the distributional properties of the speech sounds. In a seminal article, Maye et al. (2002) found that infants as early as 6 months can extract distributional regularities quickly when presented with sounds from a /da/ - /ta/ continuum. In the familiarization, infants were presented with a unimodal or a bimodal distribution, based on the frequency of each token in the continuum, resulting in one or two frequency peaks. In the test, infants only discriminated the stimuli when they had been previously familiarized with a bimodal distribution. The earliest evidence of sensitivity to distributional learning has been found with 2-3-month-old Dutch infants with vowels (Wanrooij et al., 2014). Infants were trained with either a unimodal or a bimodal /ae/ - /e/ distribution of the English vowel contrast and tested in an oddball paradigm using EEG. The participants showed a greater mismatch response after the bimodal exposure. The relevance of distributional learning has received extensive support not only from behavioral studies, but also neurophysiological data as well as computational modelling (Adriaans & Swingley, 2012; McMurray et al., 2009; Schatz et al., 2021; Vallabha et al., 2007, among others).

The critical observation supporting the existence of perceptual narrowing is that at an early age infants show discrimination abilities of a certain phonetic contrast not present in the infants' environment and that at a later age (a few months later) such discrimination abilities are no longer observed. The fact that some phonetic contrasts show a gradual decline at a group level has been attributed primarily to the existence of individual differences, with some infants showing earlier (or later) narrowing than others. Frequency of appearance of a phoneme in the native language also plays a role in the shape of perceptual narrowing (Anderson et al., 2003). For instance, infants become worse at discriminating a coronal, non-native pair of sounds (such as /ʈɜ/-/ [t̪] for English-speaking infants) than a dorsal one (such as /k'ɜ/-/q'ɜ/), due to the former being more frequent than the latter (Tobias, 1959). Given the assumption that phoneme categories emerge as the consequence of computations over a protracted period of time and a considerable amount of input, it is reasonable to assume that before perceptual narrowing effects can be observed, infants may possess some information about the distributional properties of the sounds of their language. In the present investigation we provide evidence of early representation of native language-specific knowledge of the vocalic space before discrimination starts to decline.

Different studies have shown that infants prefer to listen to lists of words that follow the properties of their native language. Jusczyk et al. (1993a) showed that 9-month-olds, but not 6 month-olds, prefer listening to lists of words that contain phonemes or combinations of phonemes only found in their native language (Experiment 1, Experiment 3), or conforming only to the phonotactics of the native language (Experiment 4) using phonemes found in both languages, i.e. English and Dutch (see also Friederici & Wessels, 1993; Jusczyk et al., 1993b; Jusczyk et al., 1994; Sebastián-Gallés & Bosch, 2002 among others for similar patterns of preference). Here we will test if 4–5-month-olds show a preference for lists of words that follow the distribution of vowels of their native language. To this end, we capitalize on the significant differences in the distribution of vowels in Catalan and Spanish.

Catalan and Spanish[18] are two Romance languages that present quite different distributions of the vocalic sounds. Spanish has a 5-vowel system, whereas Catalan has 8 vowels (Carbonell & Llisterri, 1992; Martínez-Celdrán et al., 2003). The main phonological distinction between the two is that Catalan has vowel reduction; meaning that in unstressed positions /e/, /ɛ/, /a/ are reduced to a /ə/ and /o/ and /ɔ/ are reduced to a /u/, whereas /i/ and /u/ can be found in both stressed and unstressed positions. The use of vowel reduction only in Catalan results in conspicuous differences in the vowel distribution of the two languages. Because mid vowels can only occur in stressed syllables, they are relatively infrequent: only 16% of all vowels in Catalan are mid vowels (e-ɛ, o-ɔ), while they are almost twice more frequent in Spanish (49%). Conversely, central-low vowels are twice more frequent in Catalan (61%) than in Spanish (29%). Figure 4.1. shows the specific frequency of appearance of each vowel (see Zacharaki & Sebastian-Galles, 2021, for a more complete description of the phonological properties of these two languages, see also Albareda-Castellot et al., 2011). If 4–5-month-old infants are sensitive to the distributional properties of the vowel sounds of their language of exposure, we expect to find a preference for lists of words instantiating the native pattern.



[18] Unless specified, when Catalan is mentioned, we refer to the Eastern dialect, spoken in the metropolitan area of Barcelona. As for Spanish, we always refer to the Standard Castilian one.

Figure 4.1. Distribution (in percentage) of Catalan and Spanish vowels. The height of the tongue during articulation (F1) is color coded. High vowels appear in blue, mid vowels in red and low vowels in green. The placement of the vowels in the pie charts approaches the one in the vowel space (F1/F2).

4.3. Methods

The method, data analysis and criteria for exclusion of participants were pre-registered on the OSF (Open Science Framework) database before analyzing data (Zacharaki & Sebastian-Galles, 2020, https://doi.org/10.17605/OSF.IO/2VR85).

4.3.1. Participants

Fifty 4.5-month-old monolingual infants participated in this experiment. All participants were healthy and full-term (>37 gestation weeks) coming from either a Catalan (n=25, range= 123-169 days, M=147, SD=15, 13 female) or a Spanish (n=25, range= 121-166 days, M=141, SD=13, 13 female) speaking family. The linguistic profile was calculated using an adapted version of a language questionnaire (Bosch & Sebastian-Galles, 2001). If a participant had less than 85% exposure to the dominant language, they were excluded from the final sample. Participants were recruited from private hospitals found in Barcelona (Spain) or online. Twenty-three additional infants were tested but not included in the final sample for the following reasons: not enough data (n=5), fussy/cried (n=5), failure to calibrate (n=3), bilinguals (n=3), low weight at birth (n=3), technical error (n=3), health issues (n=1). The research reported in this article was conducted in accordance with the principles expressed in the Declaration of Helsinki and approved by the local ethical committee (the clinical research ethical committee of the Parc de la Salut Mar).

4.3.2. Stimuli

The stimuli were 240 nonsense words which were organized in 10 lists for each language Catalan and Spanish respectively, i.e. 20 lists in total. Ten of the words in each list were disyllabic and two were trisyllabic. A female bilingual Catalan-Spanish speaker, highly proficient in the two languages, was recorded. The speaker is a musician, and she was taught phonetics in school, which enabled her to articulate the words fittingly. The speaker had no detectable accent in either language. The stimuli were recorded in a soundproof room in the 'Laboratori de Recerca en Infancia' at the University Pompeu Fabra using an Audio-Tecnica microphone (AT2050) at a sampling rate of 44100 Hz. We used Audacity® recording and editing software (Mazzoni, 1999) to record the stimuli and Praat (Boersma & Weenink, 2019) to extract the words from the recording and analyze their pitch and frequency. We made sure that our stimuli were not different in terms of pitch nor intensity (see table 1.) The speaker was instructed to say the words in an infant-directed manner and always to stress the penultimate syllable in each word.

Table 1. Acoustic properties of the lists and percentage of each vowel category (F1), standard deviation is inside the parentheses

	Pitch			Duration	% Vowels		
	Mean (Hz)	min	max	mean	low	mid	high
Catalan	262 (16.9)	221	301	0.84 (0.1)	61	16	23
Spanish	265 (16)	222	304	0.86 (0.1)	29	49	22

Each list mimicked the vowel distribution of either Catalan (Rafel, 1980) or Spanish (Alcina & Blecua, 1975)(see table 1 and figure 4.1.). The same **'CV**CV/ CV**'CV**CV structure was used in the two languages to ensure that the lists differed only in terms of their vowels (both structures are very common in Catalan and Spanish). The consonantal structure across the two languages was the same, the only difference was the frequency and placement of the vowels (table 2). We used Wuggy (Keuleers & Brysbaert, 2010), a multilingual pseudoword generator, to create phonotactically legal structures in Spanish. The same consonantal structure was used across the two languages, and all the words were also phonotactically permissible in Catalan. The speaker was instructed to use only the five common vowels of the two languages /a,e,i,o,u/ for all of the lists and avoid the Catalan vowels /ɛ, ə, ɔ/, so as to eliminate providing additional phonetic cues to the participants. The order of the words within each list was randomized across participants. After each word, a silence was added to reach two seconds in total.

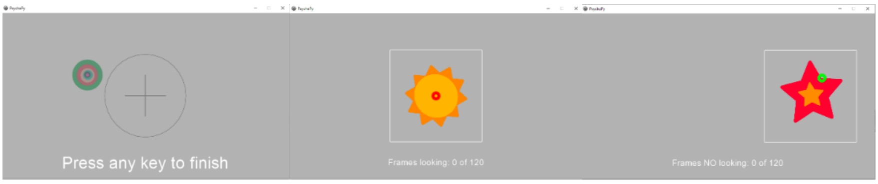
Table 2. Examples of an experimental list

Spanish	***Catalan***
/ˈbane/	/ˈbuna/
/geˈɾota/	/giˈɾota/
/ˈgiðo/	/ˈguða/
/ˈgira/	/ˈgari/
/ˈlape/	/ˈlipa/
/loˈtiɣe/	/laˈteɣa/
/ˈmaβo/	/ˈmaβa/
/ˈneβa/	/ˈnaβa/
/ˈnoɲe/	/ˈnoɲa/
/ˈpoβa /	/ˈpiβa/
/ˈreði/	/ˈraða/
/ˈtuki/	/ˈtaka/

4.3.3. Procedure

Infants were tested in a sound-attenuated, dimly lit room in the ‘Laboratori de Recerca en Infancia’ at the University Pompeu Fabra. The testing room was equipped with an ASUS VE276N (size: 27", resolution: 1920x1080) and a Tobii Pro Spectrum eyetracker (120Hz sampling rate). Two M-Audio AV 30 loudspeakers were placed left and right from the monitor, hidden behind a beige curtain. A SONY HC9 camera was placed 20 cm above the monitor allowing the experimenter to check on the infant while testing. The participants were seated on a baby chair facing the monitor at a 65 cm distance. The baby chair was adaptable, height and distance were adjusted for each participant depending on their size. The caregiver was seated in a chair behind the participant and was instructed to avoid any contact with the infant and to look towards the floor or to have their eyes closed. The experimenter controlled the experiment outside the testing room using a custom-made script on PsychoPy (Peirce et al., 2019).

We adapted the procedure employed in Jusczyk et al. (1993b), which was originally used to test infants' preference towards their predominant stress pattern. We used an eye tracker to make the procedure completely infant-controlled. Areas of interest (AOIs) were defined as enlarged squares around the visual stimuli, 600 x 600 pixels in size (see figure 4.2. B and C). The size of the AOI was 17% in relation to the whole screen size. After the infant was seated in the baby chair, a circle with a cross in the middle and a random moving ball appeared on the screen. Once the participant looked at the screen, another circle appeared that had to be green and to overlap with the preexisting circle. If the participant was not placed properly, too far from the screen or lower/higher than it should, the circle was red. In this manner, we got immediate feedback on the positioning of the participant and corrected when needed before performing the calibration (see figure 4.2. A). When the participant was placed suitably, a 5-point calibration was performed. If the participant reached the minimum of 3 valid points of calibration, a rotating sun appeared in the middle of the screen accompanied by a recurring ping-like sound. Once the participants had looked at the sun for two accumulative seconds, it disappeared, and a rotating star appeared on either side of the screen and one of the lists was played simultaneously. Each trial consisted of the sun (attention-getter) and star (visual attractor) sequence. Infants completed a practice phase with four lists (two in each language), followed by a testing phase with a maximum of sixteen trials (eight in Catalan and eight in Spanish). The auditory stimuli were presented at 65 dBs. The ordering of the lists was pseudorandomized, with the restrictions that the first list in the practice and test phases were in the native language of the participant and that no more than three lists of the same language could appear sequentially. If the participant looked outside the predefined AOI for two consecutive seconds, the trial ended. The experiment ended if the participant completed all the test trials or if they became fussy or stopped looking at the screen. The experiment lasted approximately 7 minutes. Caregivers signed a consent form before the experiment and a small gift was given to them at the end of the session.



A. **B.** **C.**

Figure 4.2. ***A.*** Circle shown before calibration as feedback for the participant's position, ***B.*** attention getter shown at the beginning of each trial, ***C.*** visual attractor while listening to the lists of words. Only the colored pictures were shown inside the testing room, the white squares were only visible in the experimenter's monitor.

4.3.4. Data analysis

Looking times were calculated using the eyetrackingR package (Dink & Ferguson, 2015) in RStudio (RStudio Team, 2019), report (Makowski et al.,2020) and a custom-made script. Looking times were calculated only when the participants were looking inside the AOIs. Looking anywhere else was considered as looking away time. For a trial to be valid, the minimum looking time was 1.25 seconds that corresponds to the longest duration of our nonsense words. For a participant to be valid, they had to have at least 4 valid trials in the test phase and a balanced number in each condition, at least two for the native condition and two for the non-native. If participants had more than 4 valid trials, they had to have an approximately equal number of trials per condition (±2).

Looking times were analyzed with linear mixed models, using *lme4* (Bates et al., 2015). The dependent variable was looking time in seconds during the test phase. The fixed effects were stimuli (Native vs. Non-native, coded as -0.5 and 0.5) and Language Profile (Catalan vs. Spanish, coded as -0.5 and 0.5). Participant number was added as a random effect. The fit of the models was evaluated using the 'anova' function. The pie plot (figure 4.1.) was generated using 'ggplot2' (Wickham, 2016) and the violin plots (figure 4.3.) using the R package 'ggstatsplot' (Patil, 2021).

4.4. Results

Mean looking times to the two types of stimuli, i.e. native and non-native, are shown in figure 4.3. We fitted a linear mixed model (estimated using REML and BOBYQA optimizer) to predict Looking Times with Stimuli as a fixed effect (formula: AOI_sec ~ Stimuli). The model included Participant as random effect (formula: ~1 | Participant.Number). The model's total explanatory power is moderate (conditional R2 = 0.19), and the part related to the fixed effects alone (marginal R2) is of 0.008. The output of the model is summarized in table 3. The model did not include Language Profile (Catalan, Spanish) nor its interaction with stimuli since we did not expect them to have any effect and these additions did not improve the fit of the model.

The model's intercept when all the predictors are centered at 0, is at 7.71 (95% CI [6.88, 8.54], t (619) = 18.25, p < .001), meaning that participants looked at the lists of words on average for 7.7 seconds. We found that the effect of Stimuli is statistically significant and negative (beta = -1.06, 95% CI [-1.89, -0.22], t (619) = -2.48, p = 0.013). Participants paid attention to the native lists 1.06 seconds more than to the non-native ones. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using the Wald approximation.

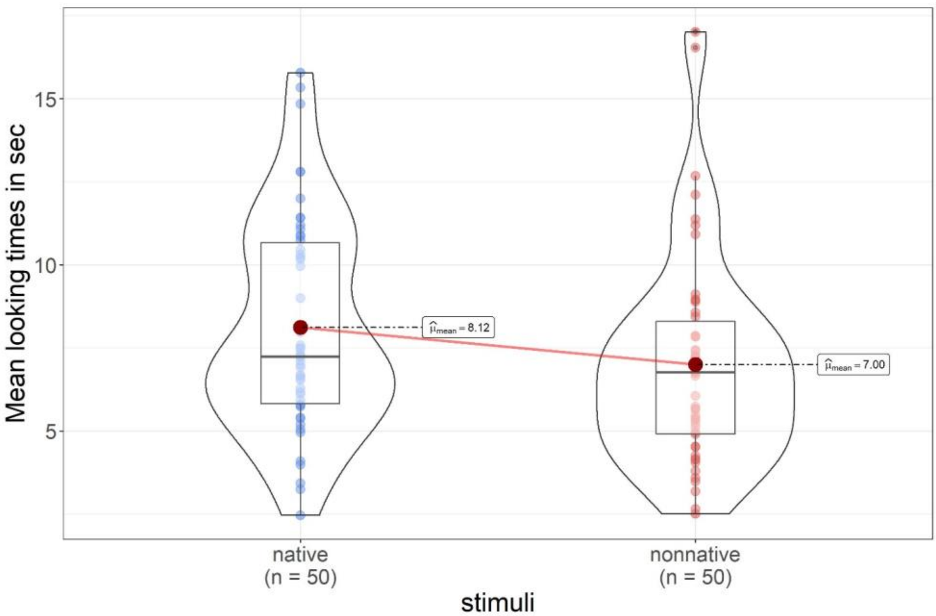


Figure. 4.3. Dots represent the individual participants’ mean looking times to the native and non-native lists. Red large dots stand for the group mean, boxplots stand for the interquartile difference, and horizontal lines for the median.

Table 3. Output of the model

	AOI_sec		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	7.71	6.88 – 8.54	**<0.001**
stimuli1	-1.06	-1.89 – -0.22	**0.013**
Random Effects			
σ^2	28.01		
τ_{00} Participant.Number	6.45		
ICC	0.19		
N Participant.Number	50		
Observations	623		
Marginal R^2 / Conditional R^2	0.008 / 0.194		

4.5. Discussion

The objective of the present study was to decipher whether infants show sensitivity to the native language vowel system around the 4th month of life, an earlier time than first signs of perceptual narrowing for vowels, i.e. 6 months, have been reported (Kuhl et al., 1992; Polka & Werker, 1994). We ran a preference procedure using the contrasting naturally occurring frequency distribution of vowels found in Spanish and Catalan. The results showed that infants prefer listening to lists that resemble the incidence of vowels in their native language over lists that do not.

Previous research investigating the establishment of the native phoneme system has compared infants’ responses to native and non-native contrasts at different points in time. The fact that infants can show discrimination at an early age, and no longer a few months later has been interpreted as indicating that the acquisition of the native phoneme system started at some point between the two testing times. The result of our investigation suggests that such approaches may have underestimated infants' linguistic knowledge. However, our results do not allow to conclude if perceptual narrowing, measured as the decline of discrimination to non-native contrasts has already started at the age we have tested. As mentioned, Bosch & Sebastián-Gallés (2003) showed that Catalan and Spanish 4.5-month-olds, the same population tested here, were equally able to discriminate the /e/ – /ɛ/ contrast, native and non-native respectively, however, at 8 months of age, only infants learning Catalan continued to discriminate the contrast. The authors concluded that perceptual narrowing to the native vowel space had not started at 4.5 months. It is possible that behavioral paradigms based on procedures measuring recovery of attention after habituation/familiarization exposures may not provide sufficiently fine-grained measures to detect the onset of the characteristic decline of non-native contrasts in perceptual narrowing. Electrophysiological measures might offer an alternative approach, as some adult studies have reported changes in neurophysiological responses preceding changes in behavior (Tremblay et al., 1998) and the same possibility has been postulated for infants (Cheour et al., 1998). However, the difficulty of getting sufficiently robust measures at the individual level, together with the individual differences in the time-course of perceptual narrowing, may make this approach methodologically quite challenging.

We have assumed that infants have used the distribution of vowels in the lists to prefer one type of list over the other. It might be argued that infants used other kinds of information also present in the stimuli. Infants could have used information about specific phonemes, however the lists in the two languages used the same consonants and vowels, so infants could not have used such information to prefer one list over another. A second type of information refers to specific properties of some words. Although the phonemes were the same in both lists, because of the difference in phonotactic rules between Catalan and Spanish, on average half of the words in each list were only permissible in Spanish, while the other half were permissible in both languages. The fact that Catalan has vowel reduction implies that only one mid-vowel can appear in a word, that is, the stimulus /go'bete/, cannot be a word in Catalan, since /o/ and /e/ are reduced when not stressed in Catalan (it should be /gu'betə/; Toro et al., 2011, showed that Catalan adults, but not Spanish ones, use this type of information to segment words in continuous speech). It is highly unlikely that infants may have used this information to prefer one type of lists over the other. First because the effect should have been present only in Catalan-learning infants (for Spanish-learning infants, all stimuli are phonotactically permissible words in their language of exposure), and our results did not show an interaction between stimuli and language of exposure. Second, because the emergence of preference towards lists of words that contain more frequent phonotactic patterns has not been reported before 9-10 months of age (Gonzalez-Gomez & Nazzi, 2015; Jusczyk et al., 1993; Jusczyk et al., 1994; Nazzi et al., 2009). A third possibility is that infants may have paid more attention to stressed than unstressed syllables. This is also unlikely because such bias would make the differences disappear between the two languages, as vowel reduction only applies to unstressed syllables, therefore no preference should be evident. Finally, a fourth type of information that infants could have paid attention to are word edges. Infants might have paid more attention to the first syllable (primacy effect) or the last one (recency effect) in each pseudoword. This effect has been shown previously with 3-month-old infants when presented with speech sequences of five syllables (Hochmann et al., 2016). The vowel information gathered from only the first or last syllable from our stimuli coincides with naturally occurring vowel distribution in the whole language, therefore we cannot be certain on if infants computed distributions over all the stimuli or just the edges. This question remains to be answered in future experiments, specifically designed to answer this issue.

We propose that before the emergence of the phonetic system, consequence of perceptual narrowing, infants compute the frequency distribution of sounds and are sensitive to speech mirroring the properties of the native language. Such a process would take place over months, likely starting prenatally. Moon et al. (2013) showed that newborns prefer listening to vowels they had been exposed to prenatally (native language) over phonemes they have never been exposed to before (foreign language), so sensitivity to speech sounds frequency can be detected already at birth. We also propose that this type of representations would contribute to the more refined language discrimination abilities (beyond the use of rhythmic/prosodic information) observed in infants around 4-5 months of age. Zacharaki & Sebastian-Galles (2021) recently reported that 4–5-month-old infants can discriminate between two dialects of Catalan only differing in the distributional properties of their vocalic systems.

Our results do not provide information on the nature of information infants compute distributional information over. There is no doubt that frequency of appearance is an important factor of native phoneme category formation (Cristia et al., 2011; Jusczyk, 1993) and in the decline of perception non-native phonemes (Anderson et al., 2003). Infants acquire first the most frequent native phonemes, such as coronal stops in English, than less frequent ones, such as dorsal stops, and consequently lose the ability to discriminate between the more frequent non-native sounds first (Anderson et al., 2003).

This pattern of behavior has been found with consonants at 8-9 months of age, earlier than the age typically associated with perceptual narrowing of consonants, i.e. 10-12 months of age (Werker & Tees, 1984). These earlier effects of frequency on speech perception align with our hypothesis and results.

In conclusion, the present study provides evidence of emergence of native phonetic knowledge at 4.5 months of age, earlier than previously reported. Infants keep track of the most frequent sounds present in their native language and they can use such knowledge when presented with novel stimuli. Future research can shed light on the exact information that infants compute their statistics over.

4.6. References

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 72–77.

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *14*(2), 395–401.

https://doi.org/10.1111/j.1467-7687.2010.00989.x
Alcina, F. J., & Blecua, J. M. (1975). Gramática española. *Barcelona: Ariel*, *1991*.
Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical Basis for Speech Sound Discrimination. *Language and Speech*, *46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601
Aslin, R., Jusczyk, P., & Pisoni, D. (1998). *Speech and auditory processing during infancy: Constraints on and precursors to language.*
Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). *Fitting Linear Mixed-Effects Models Using lme4*. *1*. https://doi.org/10.18637/jss.v067.i01
Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. http://www.praat.org/
Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189–221. https://doi.org/10.1080/09541446.2000.10590222
Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49.
Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801
Carbonell, J. F., & Llisterri, J. (1992). Catalan. *Journal of the International Phonetic Association*, *22*((1-2)), 53–56. https://doi.org/10.1017/S0025100300004618
Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. https://doi.org/10.1016/0378-5955(94)00164-L
Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561
Cristia, A., Mcguire, G. L., Seidl, A., & Francis, A. L. (2011). *Effects of the distribution of acoustic cues on infants ' perception of sibilants*. *39*, 388–402.

https://doi.org/10.1016/j.wocn.2011.02.004
Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R Library for Eye-tracking Data Analysis*.
Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303
Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287–295. https://doi.org/10.3758/BF03205263
Gonzalez-Gomez, N., & Nazzi, T. (2015). Constraints on statistical computations at 10 months of age: The use of phonological features. *Developmental Science*, *18*(6), 864–876. https://doi.org/10.1111/desc.12279
Hochmann, J. R., Langus, A., & Mehler, J. (2016). An Advantage for Perceptual Edges in Young Infants' Memory for Speech. *Language Learning*, *66*(September), 13–28. https://doi.org/10.1111/lang.12202
Jusczyk, P W, Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993a). Infants′ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/https://doi.org/10.1006/jmla.1993.1022
Jusczyk, P. W. (1993). From general to language-specific capacities: the WRAPSA Model of how speech perception develops. *Journal of Phonetics*, *21*(1–2), 3–28. https://doi.org/10.1016/s0095-4470(19)31319-1
Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993b). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development*, *64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x
Jusczyk, P. W, Luce, P. A., & Charles-Luce, J. (1994). Infants' Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, 630–645.
Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, R. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5), 606–608.

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255–259. https://doi.org/10.1017/S0025100303001373

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-11. https://doi.org/10.1016/s0010-0277(01)00157-3

Mazzoni, D. (1999). *Audacity*®.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica (Oslo, Norway : 1992)*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, *126*(3), 1440–1446. https://doi.org/10.1121/1.3158931

Pallier, C., Bosch, L., & Sebastian-Galles, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. https://doi.org/https://doi.org/10.1016/S0010-0277(97)00030-9

Patil, I. (2021). Visualizations with statistical details: The “ggstatsplot” approach. *Journal of Open Source Software*, *6*(61), 3167. https://doi.org/10.21105/joss.03167

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans XXIII*, *2*, 473–496.

RStudio Team. (2019). *RStudio: Integrated Development Environment for R RStudio, Inc.* http://www.rstudio.com/

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *PNAS*, *118*(7), 2001844118. https://doi.org/10.1073/pnas.2001844118/-/DCSupplemental.y

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 28, Issue 4, pp. 974–989). American Psychological Association. https://doi.org/10.1037/0096-1523.28.4.974

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111–123. https://doi.org/10.1016/S0010-0277(99)00024-4

Tobias, J. (1959). Relative Occurrence of Phonemes in American English. *Journal of the Acoustical Society of America*, *31*, 631.

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, *64*(2), 171–180. https://doi.org/10.1016/j.jml.2010.11.005

Trehub, S. E. (1976). *The Discrimination of Foreign Speech Contrasts by Infants and Adults*. *47*(2), 466–472.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *NeuroReport*, *9*(16). https://journals.lww.com/neuroreport/Fulltext/1998/11160/The_time_course_of_auditory_perceptual_learning_.3.aspx

Tsuji, S., & Cristia, A. (2013). *Perceptual Attunement in Vowels : A Meta-Analysis*. https://doi.org/10.1002/dev.21179

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*(FEB).

https://doi.org/10.3389/fpsyg.2014.00077
Werker, J F, & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.
Werker, J. F, & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In *The Oxford handbook of developmental psychology* (pp. 909–925).
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org
Yeung, H. H., Chen, K. H., & Werker, J. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*, 123–139.
Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. *Cognition*, *February*, 104628. https://doi.org/10.1016/j.cognition.2021.104628
Zacharaki, K., & Sebastian-Galles, N. (2020). *Early vowel preference of Catalan/Spanish 4-5-month-old infants: a distributional account.* https://doi.org/10.17605/OSF.IO/2VR85

5. Adults’ (in)sensitivity to vowel distribution information

Konstantina Zacharaki[*], Nuria Sebastian-Galles[*]

[*] Center for Brain and Cognition (CBC), Universitat Pompeu Fabra, Barcelona, Spain

5.1. Abstract

Previous linguistic knowledge modulates speech perception. One type of such knowledge is the permissible combinations of sounds into meaningful units in a language. Word recognition and speech segmentation are affected by phonotactic restrictions. However, these are localized phenomena. We investigated if adults are able to extract phonotactic-related regularities above word level, when presented with lists of words that mimic, or not, the vowel distribution of their native language. Previous research with 4.5-month-old infants, before word-level phonotactics has been reported, has shown that they are sensitive to this type of cue. The present results uncovered a different pattern of results as adults seem to rely on word-level combinations rather than list-level information.

5.2. Introduction

Languages are characterized by different phonetic repertoires which form words abiding to specific constraints on the combinations of sounds into legal structures. These permissibility rules are known as phonotactics. Humans acquire this knowledge early on and previous research has shown that it determines speech perception. For instance, English native speakers can easily tell that “katakana” is unlikely to be a word in English. If asked they might say that it is too long, that the same vowel appears too many times or even that the vowel appears at the wrong place. Adult speakers possess different types of knowledge about how a prototypical word of their native language should sound. Such knowledge may refer to specific phonemes, or suprasegmental information, like stress patterns, word length or even the distribution of phonemes within the word, i.e. phonotactics. In the present investigation we want to explore if adults can use one type of phonotactic information, the distribution of vowels appearing in lists of words, to discriminate between lists reflecting distributions of different languages.

Most studies investigating listeners' phonotactic knowledge have taken place in the context of models of word recognition. Previous research has shown that highly frequent words are recognized faster than words of low frequency (Howes, 1957). Interestingly, disyllabic nonwords are evaluated more often as phonologically 'good', i.e. sounding more like the native language of the participants, when they are made up of frequent combinations of sounds (Vitevitch et al., 1997). The contribution of phonotactic knowledge has also been studied in relation to speech segmentation. In these studies participants are usually tested using the paradigm on artificial language learning from Saffran et al., (1996). In the seminal paper, 8-month-old infants were presented with a continuous speech stream for 2 minutes, which contained four trisyllabic nonsense words. Syllables that belonged to the same word had a higher transitional probability (TP), i.e. co-occurrence frequency, than syllables expanding beyond a word's boundaries. Infants were subsequently presented with two of the words that they had heard in the speech stream and two new combinations of syllables into words. The new words could be a random combination of syllables from the familiarization, non-words, or the last syllable from a word paired with two syllables from the beginning of another word, part-word. Finn & Hudson Kam (2008) found that American adults did not segment the words they heard during familiarization when they included an illegal consonantal cluster in word initial position (i.e. a combination of consonants not appearing at the beginning of English words) but succeeded in a control experiment where words began with a legal (possible) onset. Toro et al. (2011) showed that Catalan natives applied native language's vowel reduction (i.e. no two mid vowels can appear in a word) when segmenting words from a continuous speech stream. Sensitivity to language-specific phonotactic information has been shown in a variety of languages and materials (see (Mersad & Nazzi, 2011) or (Onnis et al., 2005) among others). The results of these studies show that participants are better at identifying the embedded words when they follow the phonotactics of their native language as compared when they clash with them.

A few studies have investigated sensitivity to phonotactic information in preverbal infants. In such studies, infants are typically presented with two different types of lists of stimuli, one of them following the phonotactics of the language of exposure and the other not. These investigations have shown that young infants prefer listening to linguistic information that is consistent with distributional properties of their language of exposure: at 9 months of age infants prefer listening to lists of words that follow the predominant stress pattern of their native language (Jusczyk et al., 1993b) , at 9 months, they prefer listening to lists of words that contain highly frequent phonotactic patterns over non frequent ones and also lists that contain legal, vs illegal, consonantal clusters in word initial and word final position (Friederici & Wessels, 1993; Jusczyk et al., 1993a; Jusczyk et al., 1994), at 10 months, monolingual Catalan infants prefer listening to lists with nonwords with a legal consonantal cluster at word-final position, whereas Spanish monolingual infants showed no preference given that all stimuli were phonotactically illegal for them, since words cannot end with consonantal clusters in Spanish (Sebastián-Gallés & Bosch, 2002). Recently we have reported that such preferences can be observed for the global distribution of vowels in the native language: 4–5-month-old infants prefer listening to lists of words that resemble the native language over lists of words with a different distribution of vowels (Zacharaki & Sebastian-Galles, under review). Phonotactic knowledge has also been studied in the context of word segmentation, showing that 9-month-old infants also use probabilistic phonotactic cues to divide speech into word units (Mattys & Jusczyk, 2001).

As described, adult and infant research on sensitivity to language-specific patterns has followed two different approaches. On the one hand, focusing on the global/list level, as in the preference aforementioned studies with preverbal infants. On the other hand, focusing on local combinations, like in most adult word segmentation or recognition tasks. In the present research we want to test what type of information adults are more sensitive to, when the two types of information are present and induce different patterns of responses. We have capitalized on a phonological difference between Catalan and Spanish that results in them having a rather different distribution of vowels at the global (list) level and at the local (word) level. Spanish and Catalan are two Romance languages similar in terms of rhythm but they differ in terms of number of vowel categories and their respective frequency. Catalan has eight vowels /i,e,ɛ,a,ə,u,o,ɔ/ and Spanish only five /i,e,a,u,o/ (Carbonell & Llisterri, 1992; Martínez-Celdrán et al., 2003). As aforementioned, Catalan bears vowel reduction meaning that low and mid vowels can only appear in stressed positions, otherwise they are reduced (see figure 5.1. (Wheeler & Wheeler, 2005). The consequences of this Catalan-specific phonological property are twofold. Firstly, the vowel distributions of Catalan and Spanish are noticeably different (see table 1). The majority of vowels in Catalan are low (61%), whereas the majority of vowels in Spanish are mid (49%) (Alcina & Blecua, 1975; Rafel, 1980). Secondly, some words in Spanish are phonotactically illegal (cannot occur) in Catalan because they contain more than one mid or low vowel such as the word metro (Spanish: /ˈmɛtɾo/, Catalan: /ˈmɛtɾu/). These discrepancies between the two languages will allow us to disentangle if adults weigh stronger phonotactic regularities at the word level or also at a global, list level. We presented Catalan and Spanish adults with lists of nonwords that were representative of the vowel distribution of the two languages. Critically, on average half of the words in the lists in Spanish were phonotactically illegal in Catalan. If participants can extract the vowel distribution of the lists, we expect both Spanish and Catalan adults to identify the lists that resemble their native language. However, if participants are bound to process the lists of pseudowords locally, then we expect to find an asymmetry. Catalan participants will be able to detect the lists of words that are native-like but Spanish participants will not, given that all stimuli are phonotactically legal in their native language.

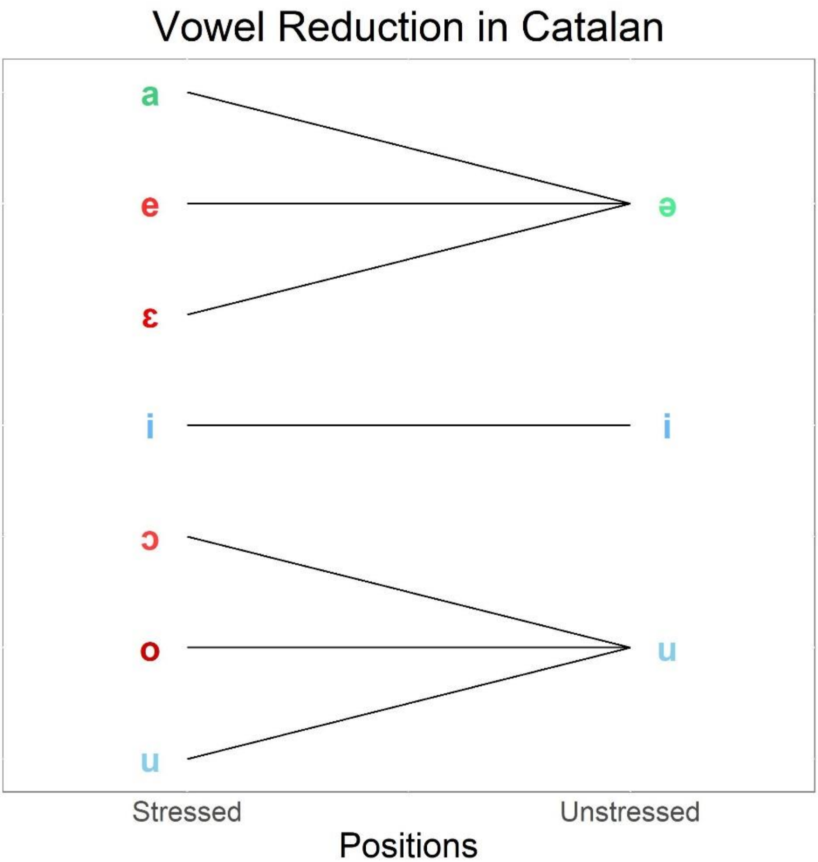


Figure 5.1. A schematic visualization of vowel reduction in Catalan. Vowels that can be used in stressed positions are plotted on the left and vowels that can be used in unstressed positions are plotted on the right. High vowels are in blue, mid vowels are in red and low vowels are in green.

Table 1. Vowel distribution (% of total vowels in the language) of Spanish and Catalan

	/a- ə /	**/e-ɛ/**	**/i/**	**/o-ɔ/**	**/u/**
Catalan (Rafel, 1980)	20+41	5+3	14	5+3	9
Spanish (Alcina & Blecua, 1975)	29	27	18	22	4

5.3. Methods

5.3.1. Participants

One hundred adult participants were tested. All participants were dominant speakers of Catalan (n=50, range=18-73, M=32, SD=16, female=37, male= 12, non-binary=1) and/or Spanish (n=50, range=18-62, M=25, SD=9, female=40, male=10). Participants' dominant language was determined as the language they chose to have the instructions of the experiment in; at the beginning of the experiment, they were given the choice to be in Catalan or in Spanish. We decided to do so because according to Grosjean (1998, 2001) the linguistic context induced by the instructions the participants received was likely to set the participants in that *language mode*, therefore making more accessible language processing mechanisms in the language of test. To assess participants' knowledge of other languages, at the end of the experiment they were asked in which languages they were fluent speakers. Four choices were provided: Catalan, Spanish, English or Other. If participants chose Catalan, another question appeared on the screen, and they had to indicate which dialect they spoke (see Appendix for more details). If they chose Other, a new blank space appeared where they had to write down the language. Out of the 50 people who participated in the experiment in Catalan, 21 did not declare to be fluent in any other language, and 29 declared to be fluent in at least another language (Spanish, English or other). Out of the 50 people who participated in the experiment in Spanish, 27 did not declare to be fluent speakers in any other language and 23 declared to be fluent in at least another language (Catalan, English or other). The percentage of coincidence between the language participants selected at the beginning of the experiment and the one(s) they declared at the end of the experiment was high (42% in the Catalan group and 78% in the Spanish group). Participants were recruited through flyers distributed at the Universitat Pompeu Fabra campus, and through emails to distribution lists of students from different Spanish Universities. The information included in the announcements made it clear that we were looking for native speakers of Catalan and Spanish only. In the analyses, we have included only participants with no history of hearing problems or learning difficulties and participants who finished the experiment. The experiment reported in this article was conducted in accordance with the principles expressed in the Declaration of Helsinki and approved by the local ethical committee, the clinical research ethical committee of the Parc de la Salut Mar. Participation was voluntary, and participants did not receive any compensation for their participation.

5.3.2. Stimuli[19]

We created 240 pseudowords organized in 20 lists in total, 10 in every language. Each list mimicked the vowel distribution of either Catalan (Rafel, 1980) or Spanish (Alcina & Blecua, 1975) (see table 1). Each list comprised of 12 words, 10 of them were disyllabic and 2 trisyllabic. The duration of each list was 24 seconds. After every pseudoword a variable silent ISI (inter stimulus interval) was added to reach 2 seconds depending on how long the pseudoword was to ensure that the stimuli were presented at a regular pace. Stress always fell on the penultimate syllable of each pseudoword (**'CV**CV, CV**'CV**CV), a common pattern in Spanish and Catalan. The same phonemes were used in the two lists. The consonantal structure was the same across the two languages, the only difference was the placement and frequency of each vowel. On average, half of the words in the Spanish lists were phonotactically illegal in Catalan. To create the pseudowords, we used a pseudoword generator named 'Wuggy' (Keuleers & Brysbaert, 2010). A female highly proficient bilingual speaker of Catalan and Spanish produced the pseudowords in an infant-directed manner. The recording took place in a soundproof room at the 'Laboratori de Recerca en Infancia' at the Universitat Pompeu Fabra using an Audio-Tecnica microphone (AT2050) at a sampling rate of 44.100 Hz. Audacity (Mazzoni, 1999) was used to record the stimuli and Praat Software to measure the acoustic features of the lists of words (Boersma & Weenink, 2019). An effort was made to ensure that the lists of words were not different in terms of pitch or fundamental frequency (see table 2).

[19] The stimuli are the same as in Zacharaki & Sebastian-Galles (submitted).

Table 2. Acoustic properties of the lists of pseudowords

	Pitch			**Word Duration**
	Mean (Hz)	min	max	mean (sec.)
Catalan	262 (16.9)	221	301	0.84 (0.1)
Spanish	265 (16)	222	304	0.86 (0.1)

5.3.3. Procedure

Participants were tested online using a custom-made Javascript allocated at the university IT services. To access the experiment, they had to click on a link (https://cbclab-online.upf.edu/sap/kzacharaki/preference_adult/) that redirected them to the experiment. At the beginning of the experiment, participants had to choose the language setting for the experiment, between Catalan or Spanish. Then, they were redirected to a screen titled 'consent form' where they were informed that we would register their responses and that participation was voluntary. We also provided the names and contact information of the authors. It was explicitly written on the screen that they could read the consent form by clicking on a hyperlinked word (here). For the experiment to start, participants had to fill in their email and tick two boxes indicating that they had read the consent form and that they gave us their consent to use their data for the purposes of the investigation. Then, a new screen appeared with the title 'Audio Adjustment'. They were instructed to adjust the volume of their device to a comfortable level by listening to a song named Alegria by Antònia Font. Under the instructions, there was a horizontal white oval shape with the play symbol (music media control). The duration of the song was also visible, 3.15 minutes in total. Participants controlled when the song was reproduced or stopped. Under this music control setup there was a horizontal bar that was by default set at 75% volume that was adjustable by the participants. Once participants had adjusted the audio levels of their device they had to click on a 'done' box to move on to the next experimental part (see figure 5.1. in the Appendix).

The instructions of the task asked the participants to imagine that they were in the future, in the year 3021, and that some archaeologists found recordings of people reciting words in two different languages. The archaeologists thought that one of the two languages was related to ancient Catalan (or Spanish, depending on whether they performed the experiment in Catalan or Spanish). Participants were told that they would listen to lists of words and their task was to evaluate whether each list was influenced by Catalan (or Spanish). Instructions only mentioned the language chosen by the participant to run the experiment. Participants were instructed explicitly to judge at the list level and to avoid paying attention to individual words. Participants answered by pressing yes or no using the keys **S** and **N** respectively or clicking with their mouse (corresponding to SI and NO in Catalan and Spanish). To minimize response bias (Kim et al., 2017) participants were asked to rate the confidence level of their response on a scale of 0 to 5: 0 stood for completely uncertain and 5 for very certain. Participants were asked to stay concentrated during the task and to avoid noisy environments. Participants heard the entire list, all 12 pseudowords in a randomized order. The first four lists were used as practice, half in Spanish and half in Catalan and participants were given visual feedback with a check (ü) or an X symbol depending on their responses. No feedback was given during the test phase. The order of the lists was pseudo-randomized, with the restriction that no more than 3 lists of the same language could appear sequentially.

Participants were given the option of explaining which type of information they used to perform the task (see Appendix). At the end of the experiment, participants were provided with feedback about their percentage of correct responses.

The data were stored in two separate files. An excel file containing the order of stimuli presentation and information on the time, day, part of the experiment, audio file reproduced and responses. A second file was a plain text document (txt.) describing the technical features of the device participants used, such as screen and memory size, and the timings of the stimuli presentation and the answers to the demographic and linguistic questionnaire. Data were stored in a safe repository at the Universitat Pompeu Fabra.

5.4. Results

Participants' binary responses (yes/no) were converted into hits and correct rejections. We calculated the A' score which is the non-parametric counterpart of d' score (Macmillan, 1993; Stanislaw & Todorov, 1999). A' ranges from 0 to 1, 0.5 corresponding to chance level and 1 to perfect performance. We used a custom script in RStudio (RStudio Team, 2019) to calculate the A' scores (Pallier, 2002) and to perform the statistical analyses.

Figure 5.2. shows the proportion of correct responses for each group of participants. We checked the homogeneity in variance of the two groups of participants using the F-test. The test indicated that the two groups' variance was unequal ($p < .001$). Independent samples t-test of unequal variance showed significant differences between the A' scores for each group (see also figure 5.3.) ($t(70.17) = 13.37$, $p < .001$, $d = 2.67$). Catalan natives performed better than Spanish ones (A'= 0.90 and A'= 0.44 for each group respectively). T-tests against chance levels showed that Catalan natives performed above chance ($t(49) = 27.63$, $p < .001$, $d = 3.91$), but Spanish natives did not ($t(49) = -1.64$, $p = .947$, $d = -0.23$). We also performed an unpaired two-samples Wilcoxon test (Mann-Whitney test) on the mean confidence ratings of participants separated by the language of test, which showed that the two groups differed in their confidence ratings ($W = 1544.5$, p-value = 0.043). Catalan natives were slightly more confident in their answers (Mdn=3.12) than Spanish natives (Mdn=2.81) (see figure 5.4.).

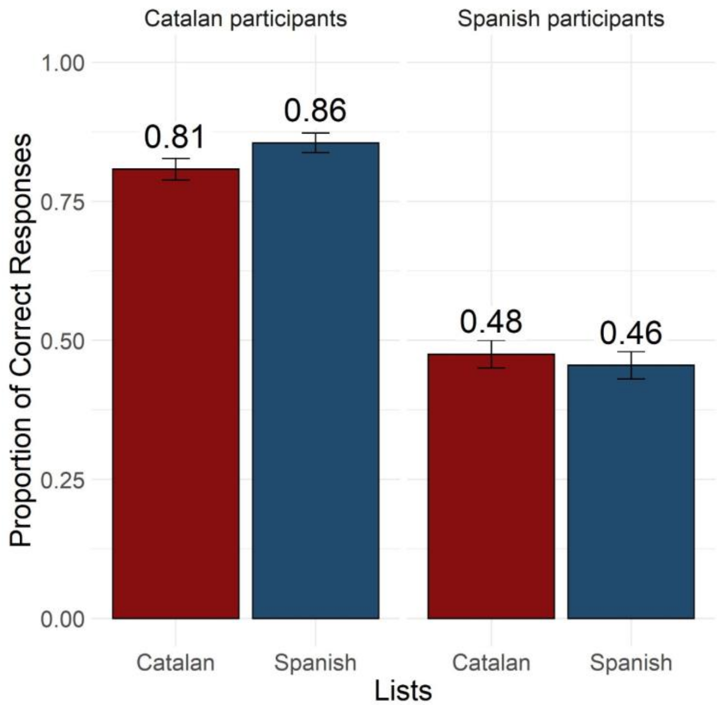


Figure. 5.2. Mean proportion of correct responses to the two types of lists are plotted separately for Catalan (dark red) and Spanish (dark blue) participants. Error bars stand for the standard error of the mean (SEM).

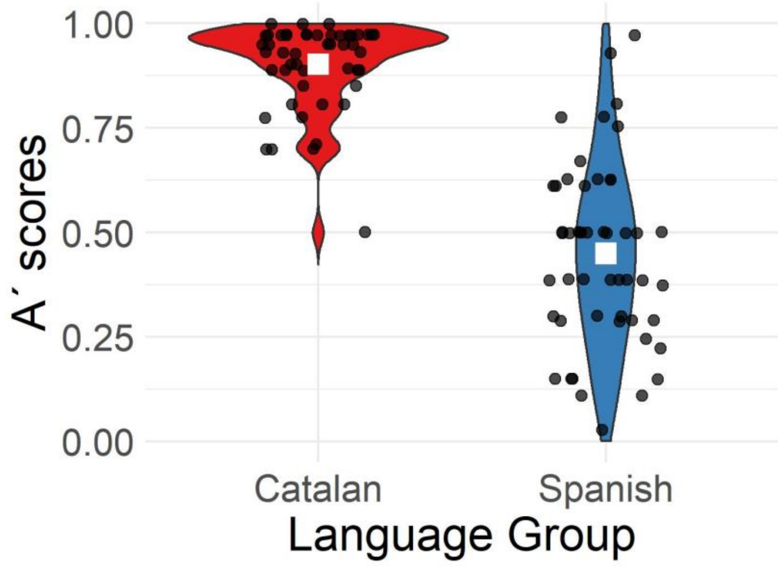
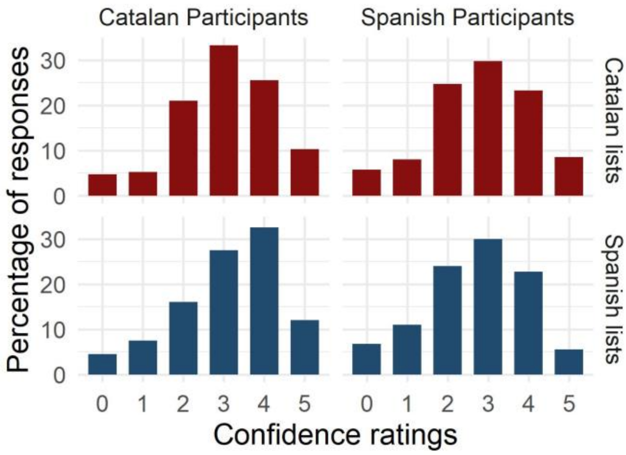


Figure 5.3. Individual A' scores plot. White squares correspond to the mean of each group and violins show the distribution of values.

A.



B.

Lang. Group	Median	IQR
Catalan	3.12	1.03
Spanish	2.81	0.98

Figure 5.4. ***A.*** Participants' confidence ratings are plotted separately for the two types of lists. The Catalan group is plotted on the left column and the Spanish group on the right column. The upper row contains ratings for the Catalan lists and the lower row contains ratings for the Spanish lists. ***B.*** Descriptive statistics (median and interquartile range) summarized in the table per language group.

5.5. Discussion

The results of the present investigation showed that only Catalan natives were able to tell the differences between the two types of lists. Spanish participants were unable to distinguish between the two types of lists, since all of the pseudowords were phonotactically legal in their native language. As expected, Spanish participants were less confident in their responses. This pattern of results supports the hypothesis that participants were paying attention to the properties of specific words and not to the overall distribution of vowels in the lists. There is very little experimental research addressing the types of information listeners use to classify speech streams as native or not. Different lines of research provide converging evidence on the importance of word-level and phonotactic information in adult language identification and word processing. Experimental evidence on the identification of dialects of Swiss German (Leemann et al., 2018), Dutch and English (Van Bezooijen & Gooskens, 1999) provide supportive evidence of the pivotal contribution of segmental information in comparison to prosodic one for adults.

A second line of research relevant to interpret our results refers to the use of phonotactic knowledge in word recognition. In the early work by Vitevitch et al. (1997) it was shown that nonwords were evaluated as being more native-like when they were composed of highly probable phonotactic structures. Likewise, repetition was more accurate and faster depending on the frequency of segments and their sequences (Vitevitch & Luce, 1998). Interestingly, the permissibility of nonwords is encoded even in passive listening tasks. German adults showed a larger N400 when they were presented with a legal pseudoword than when it was illegal regardless of register, infant or adult directed speech (Rossi et al., 2011). Adults seem to have focused and benefited from the sublexical, phonotactic patterns in such tasks. Research using real words has provided conflicting results. When participants have to repeat a word that has high neighborhood density, many words are phonetically similar, an inhibitory effect is observed, participants are slower (Vitevitch et al., 1999; Vitevitch & Luce, 1998). Additionally, adults seem to use phonotactics as a heuristic to locate word boundaries when words and phonotactic boundaries are aligned (McQueen, 1998).

One important conclusion of our research is the striking differences between the patterns of results with young infants and the adult ones. As mentioned before, using the same materials as here, Zacharaki & Sebastian-Galles (under review) observed that Catalan-learning as well as Spanish-learning 4.5-month-old infants were equivalent in their abilities of discriminating between Catalan and Spanish lists. We hypothesize that the differences between the two populations may be due to a combination of on the one hand adults' possession of a well-established phonetic system, together with better cognitive processes, in particular memory. Both dimensions would allow adults to hold more detailed representations of the presented stimuli. Different studies have shown that proficiency shapes the way information is processed: musicians perceive a melody as a sequence of related sounds, while the naïve listener focuses on the melodic contour as a total (Bever, 1975). Likewise, adults might give more weight to phonetic information (critical to comprehend language) rather than appreciate relationships at the global level.

The present results need to be replicated with future work. One of the limitations in our experiment is that the classification of participants being a Catalan, or a Spanish speaker were based on the language they chose upon testing. This is a very rough measure of language dominance and does not preclude that our participants were bilingual in both languages, at least for the Catalan group. Participants completed a short language questionnaire at the end that allowed us to check whether they are fluent in both Catalan and Spanish. Out of the 50 participants in Catalan, 29 reported being fluent in Spanish too. Out of the 50 participants in Spanish, 11 reported being fluent in Catalan too. This limitation is particularly important for the group that chose Catalan as the language of test. According to the official survey of the Catalan government concerning knowledge of languages, 99.8% of the population living in Catalonia understands Spanish and 99.7% can speak it, such percentages are in contrast with the ones collected in the present research(https://www.idescat.cat/serveis/biblioteca/docs/cat/eulp2013.pdf). It is very likely that knowledge of Spanish was underestimated in the Catalan group; given that we tested Spanish participants from outside Catalonia, the real knowledge of Catalan in the group of participants who was tested in Spanish cannot be guessed. Nevertheless, it is worth noticing that in spite of the roughness of the way language dominance was determined, the differences between the two populations are very robust. In the future, it would be fruitful to test participants of other language backgrounds to investigate whether our results are generalizable and replicable. Lastly, in our experiment we focused on the phonotactic aspect of language, by doing so we overlooked the prosodic information as we kept stress positioning constant and segmental information since we used only common phonemes from the two languages. Future experiments can attempt to see how the integration of these different levels of information may affect adults' abilities to detect phonotactic patterns.

5.6. References

Alcina, F. J., & Blecua, J. M. (1975). Gramática española. *Barcelona: Ariel, 1991*.

Bever, T. G. (1975). Cerebral Asymmetries in Humans Are Due To the Differentiation of Two Incompatible Processes: Holistic and Analytic. *Annals of the New York Academy of Sciences*, *263*(1), 251–262. https://doi.org/10.1111/j.1749-6632.1975.tb41589.x

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. http://www.praat.org/

Carbonell, J. F., & Llisterri, J. (1992). Catalan. *Journal of the International Phonetic Association*, *22*((1-2)), 53–56. https://doi.org/10.1017/S0025100300004618

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287–295. https://doi.org/10.3758/BF03205263

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149.

Grosjean, F. (2001). The bilingual's language modes' in J. Nicol(ed): One Mind, Two Languages: Bilingual Language Processing. *Oxford: Blackwell. Kehoe, MM, Lle, C. & Rakow, M.(2004). Voice Onset Time in Bilingual GermanSpanish Children. Bilingualism: Language and Cognition*, *7*(1), 7188.

Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. In *The Journal of Abnormal and Social Psychology* (Vol. 54, Issue 1, pp. 75–85). American Psychological Association. https://doi.org/10.1037/h0043830

Jusczyk, P W, Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993a). Infants′ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/https://doi.org/10.1006/jmla.1993.1022

Jusczyk, Peter W., Cutler, A., & Redanz, N. J. (1993b). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development*, *64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, Peter W, Luce, P. A., & Charles-Luce, J. (1994). Infants' Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, 630–645.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Leemann, A., Kolly, M. J., Nolan, F., & Li, Y. (2018). The role of segments and prosody in the identification of a speaker's dialect. *Journal of Phonetics*, *68*, 69–84. https://doi.org/10.1016/j.wocn.2018.02.001

Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In *A handbook for data analysis in the behavioral sciences: Methodological issues.* (pp. 21–57). Lawrence Erlbaum Associates, Inc.

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255–259. https://doi.org/10.1017/S0025100303001373

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. In *Cognition* (Vol. 78, Issue 2). https://doi.org/10.1016/S0010-0277(00)00109-8

Mazzoni, D. (1999). *Audacity®*.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. https://doi.org/10.1006/jmla.1998.2568

Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory and Cognition*, *39*(6), 1085–1093. https://doi.org/10.3758/s13421-011-0074-3

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Pallier, C. (2002). Computing discriminability and bias with the R software. *URL Http://Www. Pallier. Org/Ressources/Aprime/Aprime. …*, 1–6. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.

7769&rep=rep1&type=pdf
Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans XXIII*, *2*, 473–496.
Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience*, *23*(7), 1752–1764. https://doi.org/10.1162/jocn.2010.21547
RStudio Team. (2019). *RStudio: Integrated Development Environment for R RStudio, Inc.* http://www.rstudio.com/
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. http://www.ncbi.nlm.nih.gov/pubmed/8943209
Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 28, Issue 4, pp. 974–989). American Psychological Association. https://doi.org/10.1037/0096-1523.28.4.974
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(I), 37–149.
Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, *64*(2), 171–180. https://doi.org/10.1016/j.jml.2010.11.005
Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, *18*(1), 31–48. https://doi.org/10.1177/0261927X99018001003
Vitevitch, M S, Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, *68*(1–2), 306–311. https://doi.org/10.1006/brln.1999.2116
Vitevitch, Michael S., & Luce, P. A. (1998). When words compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064
Vitevitch, Michael S., Luce, P. A., Charles-Luce, J., & Kemmerer,

D. (1997). Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words. *Language and Speech*, *40*(1), 47–62. https://doi.org/10.1177/002383099704000103
Wheeler, M. W., & Wheeler, R. L. M. W. (2005). *The Phonology of Catalan*. OUP Oxford. https://books.google.gr/books?id=PkoVDAAAQBAJ

5.7. Appendix A.

Volume adjustment screen

Figure 5.1. Adjustment of volume screen presented to the participants before the experiment, the text is in Spanish.

Language questionnaire presented at the end of the experiment

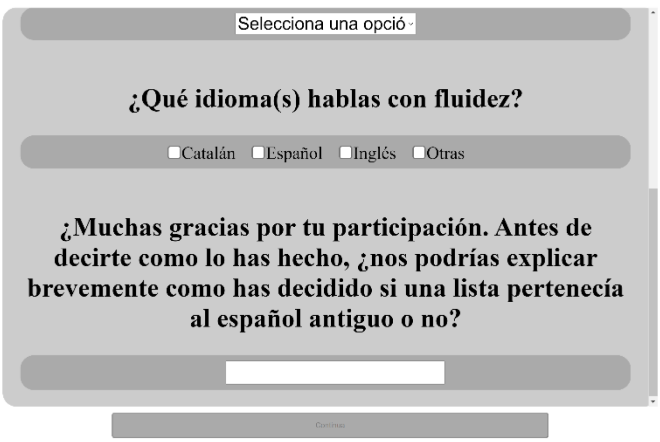


Figure 5.2. Participants were asked to choose which language(s) they speak fluently, here is the Spanish version. The choices were Catalan, Spanish, English or Other.

If participants disclosed that they were fluent in Catalan a new question appeared on the screen. Participants were asked to specify which dialect is their native one from a list of different options (see figure 5.3.). In parentheses we provided a representative list of the cities where the dialects are spoken when needed. The reason we added this clarification question is that not all Catalan dialects have vowel reduction as the one we tested here. This means that for some Catalan dialects, vowel distribution is closer to the Spanish one. Eastern Catalan, i.e. Oriental and Balearic, have vowel reduction, whereas Western Catalan, i.e. Occidental/South and Valencian do not. Our sample consisted mainly of participants who spoke the Oriental dialect of Catalan (n=37), 5 reported speaking the Occidental dialect, 4 the Balearic and the remaining 4 preferred not to answer (see table 1 for their responses).

Table 1**.** Catalan participants' responses are presented in percentages organized based on their dialect. Numbers in parentheses represent the number of participants in each group.

Dialect	**Correct (%)**	**Incorrect (%)**
Balearic (n=4)	89	11
Occidental/South (n=5)	81	19
Oriental (n=37)	82	18
N/A (n=4)	88	12

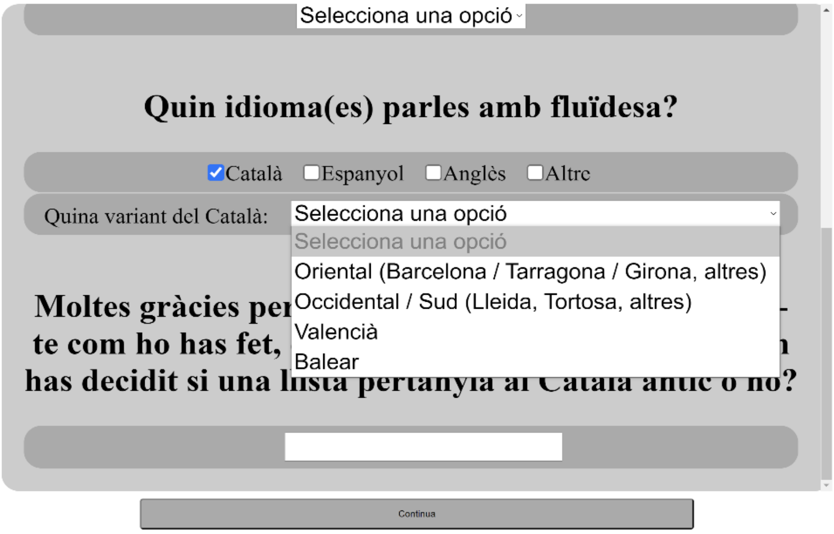


Figure 5.3. Language questionnaire showing the dialect options for Catalan presented at the end of the experiment, the text is in Catalan.

Table 2. Stimuli used in the experiment organized in the lists presented:
pseudowords in bold are phonotactically illegal in Catalan

Catalan pseudowords			Spanish pseudowords		
List num	Word	Pronunciation	List num	Word	Pronunciation
1	kadita	/ka'ðita/	1	kidote	**/ki'ðote/**
	kitxa	/'kitʃa/		kuche	**/'kutʃe/**
	lita	/'lita/		lota	/'lota/
	lona	/'lona/		lena	/'lena/
	mida	/'miða/		meda	/'meða/
	muga	/'muɣa/		mage	**/'maɣe/**
	nurra	/'nura/		narro	**/'naro/**
	patena	/pa'tena/		petina	**/pe'tina/**
	peda	/'peða/		pedi	/'peði/
	poga	/'poɣa/		pigo	**/'piɣo/**
	raka	/'raka/		roki	/'roki/
	raña	/'raɲa/		roña	/'roɲa/
2	dakena	/da'kena/	2	dukoni	/du'koni/

	gabuta	/ga'buta/		gobete	**/go'bete/**
	koba	/'koβa/		kaba	/'kaβa/
	kona	/'kona/		kani	/'kani/
	laga	/'laɣa/		lige	**/'liɣe/**
	laki	/'laki/		lake	**/'lake/**
	miga	/'miɣa/		migo	**/'migo/**
	pada	/'paða/		poda	/'poða/
	pama	/'pama/		pime	**/'pime/**
	puga	/'puɣa/		page	**/'paɣe/**
	riba	/'riβa/		rebo	**/'reβo/**
	riga	/'riɣa/		rago	**/'raɣo/**
3	diruna	/di'ɾuna/	3	dirone	**/di'rone/**
	kala	/'kala/		kela	/'kela/
	kana	/'kana/		kina	/'kina/
	marobi	/ma'roβi/		murobe	**/mu'ɾoβe/**
	marre	/'mare/		morri	/'mori/
	narra	/'nara/		narri	/'nari/
	niga	/'niɣa/		noge	**/'noɣe/**
	petxa	/'petʃa/		piche	**/'pitʃe/**
	rada	/'raða/		redo	**/'reðo/**
	rala	/'rala/		rola	/'rola/
	ruka	/'ruka/		rake	**/'rake/**
	tomi	/'tomi/		tama	/'tama/
4	bopa	/'bopa/	4	bepi	/'bepi/
	dala	/'dala/		dela	/'dela/
	gara	/'gaɾa/		geri	/'geɾi/
	kuda	/'kuða/		kode	**/'koðe/**
	liña	/'liɲa/		laña	/'laɲa/
	mepa	/'mepa/		mopa	/'mopa/
	miya	/'miɣa/		moyi	/'moʝi/
	rako	/'rako/		raki	/'raki/
	ramena	/ra'mena/		rimeno	**/ri'meno/**
	rapa	/'rapa/		rapo	**/'rapo/**
	tilura	/ti'luɾa/		telaro	**/te'laɾo/**
	titxa	/'titʃa/		tuche	**/'tutʃe/**
5	bama	/'bama/	5	bome	**/'bome/**
	daga	/'daɣa/		dagi	/'daɣi/
	dameni	/da'meni/		demano	**/de'mano/**

	dara	/'daɾa/		dora	/'doɾa/
	kotxi	/'kotʃi/		kecha	/'ketʃa/
	kupa	/'kupa/		kepi	/'kepi/
	lirota	/li'ɾota/		lorate	**/lo'ɾate/**
	luma	/'luma/		lami	/'lami/
	maki	/'maki/		moki	/'moki/
	paka	/'paka/		peki	/'peki/
	paña	/'paɲa/		pañe	**/'paɲe/**
	rega	/'reɣa/		rugo	**/'ruɣo/**
6	barri	/'bari/	6	berra	/'bera/
	biga	/'biɣa/		buga	/'buɣa/
	binaya	/'binaɣa/		bineyo	**/bi'neɟo/**
	deba	/'deβa/		dibe	**/'diβe/**
	duña	/'duɲa/		doñe	**/'doɲe/**
	keda	/'keða/		koda	/'koða/
	kona	/'kona/		kena	/'kena/
	lapa	/'lapa/		lopa	/'lopa/
	mani	/'mani/		mine	**/'mine/**
	noma	/'noma/		nemi	/'nemi/
	tada	/'taða/		tado	**/'taðo/**
	tagura	/ta'ɣuɾa/		tigaro	**/ti'gaɾo/**
7	datxa	/'datʃa/	7	diche	**/'ditʃe/**
	dota	/'dota/		dote	**/'dote/**
	gala	/'gala/		gela	/'gela/
	giba	/'giβa/		gaba	/'gaβa/
	kera	/'keɾa/		kure	**/'kuɾe/**
	lida	/'liða/		loda	/'loða/
	mupa	/'mupa/		mipa	/'mipa/
	naragi	/na'ɾaɣi/		norega	**/no'ɾeɣa/**
	neda	/'neða/		nodi	/'noði/
	parriña	/pa'riɲa/		porreñi	**/po'reɲi/**
	rotxa	/'rotʃa/		recha	/'retʃa/
	tuña	/'tuɲa/		tiño	**/'tiɲo/**
8	bada	/'baða/	8	bido	**/'biðo/**
	beña	/'beɲa/		boñe	**/'boɲe/**
	kuma	/'kuma/		kami	/'kami/
	loga	/'loɣa/		loge	**/'loɣe/**
	makana	/ma'kana/		mikone	**/mi'kone/**

	miña	/'miŋa/		miñe	**/'miɲe/**
	nala	/'nala/		nela	/'nela/
	nuga	/'nuɣa/		naga	/'naɣa/
	pina	/'pina/		pona	/'pona/
	radota	/'raðota/		reduto	**/'reðuto/**
	ritxa	/'ritʃa/		riche	**/'ritʃe/**
	tiga	/'tiɣa/		taga	/'taɣa/
9	buna	/'buna/	9	bane	**/'bane/**
	garri	/'gari/		girra	/'gira/
	girota	/gi'ɾota/		gerota	**/ge'ɾota/**
	guda	/'guða/		gido	**/'giðo/**
	latega	/la'teɣa/		lotige	**/lo'tiɣe/**
	lipa	/'lipa/		lape	**/'lape/**
	maba	/'maβa/		mabo	**/'maβo/**
	naba	/'naβa/		neba	/'neba/
	noña	/'noŋa/		noñe	**/'noɲe/**
	piba	/'piβa/		poba	/'poβa/
	rada	/'raða/		redi	/'reði/
	taka	/'taka/		tuki	/'tuki/
10	bigata	/bi'ɣata/	10	bigota	/bi'ɣota/
	daga	/'daɣa/		dige	**/'diɣe/**
	deña	/'deɲa/		diña	/'diɲa/
	gita	/'gita/		geto	**/'geto/**
	kipa	/'kipa/		kape	**/'kape/**
	latxa	/'latʃa/		locha	/'lotʃa/
	lurra	/'lura/		lorra	/'lora/
	mirre	/'mire/		mirre	**/'mire/**
	nabola	/na'βola/		nabule	**/na'βule/**
	napa	/'napa/		napo	**/'napo/**
	poka	/'poka/		peko	**/'peko/**
	tura	/'tura/		tire	**/'tire/**

Answers from the debriefing question

Table 3. Participants' answers on how they discriminated the lists organized by topic. Numbers stand for the number of participants.		
Cues used	Catalan	Spanish
intuition	3	1
phonetics	6	10
intonation	1	1
phonotactics		3
phonology		1
morphemes	1	
vowels	8	2
words	7	16
NA	24	16
SUM	50	50

6. Discussion

In the present dissertation we investigated 4.5-month-old infants' speech perception right before the first signs of perceptual narrowing, an age that has not been studied extensively. Our focus was the emergence of segmental knowledge related to the vowel distribution. To address that we tested infants in a discrimination and a preference task.

In chapter 2, we explored the ability of infants to discriminate between two pairs of languages using segmental information (or intonation). In chapter 3, we tested adults' ability to use intonational cues to discriminate Western Catalan from Spanish.

In Chapter 4, we addressed the possibility of infants having some phonemic information in place at 4.5 months of age by investigating their preference to lists of words that resemble the vowel distribution of their native language or not. In Chapter 5, we conducted research on the kind of information adults focus on, globally or locally, when presented with the same stimuli as in Chapter 4.

In this section of the thesis, I will first summarize the results from the research reported in Chapter 2 to Chapter 5. I will then discuss how the results fit in with research conducted in the field. I will also address the limitations of the experiments we ran and propose future directions. Lastly, I will conclude what has been the main contribution of this thesis.

6.1. Summary of the findings

6.1.1. Language discrimination – Infants

In Chapter 2 our goal was to evaluate whether 4.5-month-old infants have language-specific information on the distribution of vowels, an issue that has not been tested before this particular age. To assess that we ran 4 experiments using two pairs of languages. We tested a pair of languages with very different vowel distributions (Eastern and Western Catalan) and a pair of languages with comparable vowel distributions (Western Catalan and Spanish) using both natural and low-pass filtered speech.

We used the same experimental paradigm as Bosch & Sebastian-Galles (2001) to test Catalan and/or Spanish 4.5-month-old infants. We took advantage of the differences in vowel incidence naturally found in Eastern Catalan, Western Catalan and Spanish. The differences are a result of the existence of vowel reduction only in Eastern Catalan. To assess whether infants use segmental cues, we tested Catalan-dominant infants in their ability to discriminate Eastern and Western Catalan using natural sentences (Experiment 1) and low-pass filtered sentences (Experiment 3) and Spanish-dominant infants in their ability to discriminate Western Catalan from Spanish using natural stimuli (Experiment 2) and low-pass filtered stimuli (Experiment 4). The syllogism behind these experiments was that if infants were sensitive to differences in vowel distributions, they would be able to discriminate between natural sentences in Eastern and Western Catalan. While Western Catalan and Spanish should not be discriminated based on segmental cues.

Our results showed that infants can use segmental information to distinguish two languages that differ in their vowel distribution. As expected, we observed that 4.5-month-old infants could discriminate Eastern and Western Catalan in Experiment 1 (natural stimuli) but not when segmental information was reduced (Experiment 3). One question that remained open was whether infants had prior knowledge of the vowel distribution of their native language or if they extracted this information during the familiarization and testing phase of the experiment. These two different possibilities cannot be disentangled completely. We addressed this issue in Chapter 4.

We also found clear signs of discrimination in Experiment 2 and a marginal effect in Experiment 4. Thus, our findings suggest that infants used other cues, likely on intonation. The effects observed were rather weak which could be due to age constraints. Infants might need additional exposure to their language to fully acquire the native intonational cues.

In Experiment 1 and Experiment 2 we tested two groups of infants, monolinguals, and bilinguals. Based on previous findings, our hypothesis was that we would not observe any differences between the two groups. The comparison of the two populations confirmed our hypothesis.

6.1.2. Language discrimination – Adults

In Chapter 3 we wanted to investigate whether adults can use intonational cues to discriminate between prosodically similar languages. The main motivation behind this research was that previous experiments have provided conflicting results on the importance of intonational patterns in the discrimination of languages. Our goal was to extend the research on adult language discrimination using intonational differences and provide additional evidence on native and non-native speakers' discrimination capacities since there are very few available studies on the matter. In the present research, we tested the ability of adults to discriminate between Western Catalan and Spanish, two languages that are characterized by different intonational patterns. We used the same stimuli, low-pass filtered sentences, as in Experiment 4 reported in Chapter 2. We tested two groups of participants, one being native in Catalan and/or Spanish, and the other being non-native speakers of different languages. We adapted the paradigm by Ramus & Mehler (1999) to an online setup. Participants were asked to distinguish the sentences and classify them into two invented language categories. Our findings suggest that both native and non-native participants are sensitive to the intonational differences between Western Catalan and Spanish, a pair of languages that has not been tested before with adults.

6.1.3. Vowel preference – Infants

In Chapter 4, we tested whether infants have a representation of the native vowel distribution in place, prior to testing, by using a preference paradigm. The reason that prompted us to explore this issue is that our results from Chapter 2, showed that infants use segmental information to discriminate languages (Experiment 1 and Experiment 3), and this could have been observed due to the vowel information extracted during the experiment or due to prior knowledge of the native vowel space. We tested Catalan and Spanish 4.5-month-old infants on their preference to lists of nonwords that reflect the vowel distribution of their native language (either Eastern Catalan or Spanish). Critically, we did not familiarize infants to one of the two languages before testing to ensure that infants were using their previous knowledge of the native vowel system. As hypothesized, we found that infants prefer listening to the lists resembling the vowel distribution of the native language, suggesting that infants have in place some phonemic knowledge on vowels. These results also provide support that infants in the language discrimination experiments used prior phonemic knowledge.

6.1.4. Vowel Preference – Adults

In Chapter 5, we explored whether adult participants could extract information on the vowel distributions of lists of nonwords (same stimuli as in Chapter 4) that resemble their native language or not. The objective of this experiment was to investigate the level of information they would focus on. Adults can use information at a global level, here vowel distributions, or at a local level, phonotactic patterns. To tease apart these two possible heuristics, we asked adult participants to evaluate lists of nonwords as being similar or not to their native language. They were specifically instructed to pay attention to the lists and not individual words. We found that only Catalan participants succeeded in identifying the lists of nonwords that are native like. The big difference in performance between the two language groups led us to conclude that adults used local properties of speech instead of concentrating on the global, list level.

6.2. General discussion

The main contribution of the present thesis is that we provided the first evidence showing that infants have phonemic knowledge before the signs of perceptual narrowing for vowels are evident (Kuhl et al., 1992; Polka & Werker, 1994). These results have important implications on the understanding of language acquisition. Previous research has focused on the investigation of language capacities at birth (0-2 months of age) and during the second semester of life. As a result, we have a very limited understanding of what happens in between, a period during which infants evolve from universal to language-specific listeners.

Phonetic categories for vowels are not likely to emerge sharply at 6 months. The basis of distributional learning, the proposed mechanism behind phoneme acquisition, is that infants are accumulating data on the frequency of elements from their environment (Maye et al., 2002). Once infants have collected enough information and have built their phonetic categories, is when we can observe differences in native and non-native speech perception. Previous research has shown that infants as early as 2-3 months of age are susceptible to distributional information on vowels (Wanrooij et al., 2014). Additionally, the first signs of perceptual narrowing for vowels occur at 6 months of age (Kuhl et al., 1992; Polka & Werker, 1994). Thus, it appears likely that infants have some phonemic information at their disposal even before 6 months of age, most likely on vowels. To investigate whether infants have an early representation of phonemic information we ran a series of experiments.

To address whether infants have some information on the native vocalic space, we tested 4.5-month-old infants on their ability to discriminate Eastern and Western Catalan. These languages are similar in terms of rhythm and intonation, leaving only segmental cues to be a distinctive variable. The differences between Eastern and Western Catalan as aforementioned are due to vowel reduction. Our hypothesis was the infants, although they do not have phonetic categories in place yet, have a broad representation of the incidence of vowels used in their environment. To ensure that infants used this type of information we conducted an experiment using natural sentences (Experiment 1) and an additional one with low-pass filtered speech (Experiment 3). Our results supported our hypothesis since discrimination was observed only for the natural sentences, i.e. segmental information is necessary to enable discrimination.

The first way we used to tackle our hypothesis was through language discrimination. Previous work on language discrimination has failed to test the contribution of segmental cues due to a general disbelief that such knowledge exists before 6 months of age.

Numerous of the previous results observed could have been due to differences in the frequency of vowels, however it was never addressed. Of particular interest are the results from Bosch & Sebastian-Galles (1997, 2001) who tested infants' capacities to recognize or discriminate Eastern Catalan and Spanish. Infants discriminated between the two languages and attended faster to their native language. Participants could have used two kinds of information, the differences in vowel distribution or intonation. Infants recognized their native language when the stimuli were natural or low-pass filtered speech. Therefore, infants might have used both types of information. Our contribution that Bosch & Sebastian-Galles (1997) did not consider, is the importance of vowels. We found that vowel information is available and used at this age, since we only observed discrimination of natural sentences in Eastern and Western Catalan. We have provided evidence of the crucial role of segmental information for language discrimination for the first time.

However, the results of the language discrimination experiments did not allow us to determine if infants had a representation of the vowel distribution of their native language before performing the task or if they calculated it during the experiment. We tackled this issue in Chapter 4, using a paradigm with no familiarization phase. We tested infants in a preference experiment with lists of nonwords. This technique has been used successfully to investigate the emergence of phonotactic knowledge and other language specific patterns in slightly older infants (Friederici & Wessels, 1993; Jusczyk et al., 1993; Jusczyk et al., 1993). The lists contained phonotactically legal nonwords in both languages (regarding consonants) following the vowel distribution of the native language, either Eastern Catalan or Spanish. We found that infants show a preference for the lists that mimic the vowel distribution of their native language, supporting the notion that infants have some phonemic information in place before turning 6 months of age.

There are different possibilities concerning the information triggering preference. Segments were the same across the two languages, as we used only common phonemes (consonants and vowels). Therefore, infants could not have used this type of information as a cue. Infants could have used phonotactic information on the placement of vowels in each word. For instance the nonword /mi'kone/ from the Spanish stimuli is phonotactically illegal in Catalan since it is composed of two mid vowels (/o, e/). Morphologically simple words can have only one mid vowel in stressed position, namely for the nonword to be legal in Catalan it should be /mi'konə/. If infants had used phonotactic patterns in terms of vowel placement inside the words, we would expect only Catalan infants to show a preference. The stimuli for Spanish participants were all phonotactically permissible. We did not observe an asymmetry between the two groups of participants, given that adding the language profile of the participants did not improve the fit of the model. Therefore, we can also rule out the contribution of phonotactic knowledge regarding vowel placement. A third option would be that infants were paying attention to the frequency of vowels in individual nonwords (probabilistic phonotactics). Not all tokens are equally representative of the Catalan or Spanish vowel distribution. A case in point is the nonword /ma'kana/ that is more probable to be a Catalan word due to the high incidence of the low vowels (/a-ə/, 61%) in Catalan than in Spanish (/a/, 29%). In contrast, other nonwords are potential candidates of both Catalan and Spanish (/'pina/). It is highly unlikely that infants have used this kind of information as previous research has found that such knowledge emerges at 9 months and we would not have observed any preferences at the list level (Jusczyk et al., 1994; Nazzi et al., 2009). Based on our findings, we adjudicate that infants used the frequency of vowels found in each list to perform the task. Previous studies have also shown that infants can compute the frequency of phonemes very rapidly (Wanrooij et al., 2014).

Taken together our findings on the use of vowel distribution have important ramifications. We have found evidence for the first time that infants possess an early representation of the vowel space before language-specific behavior is observed. Our results complement previous theories on language perception in infancy that have highlighted the importance of vocalic information (Mehler et al., 1996). We know that infants have not attuned to the native vowel space yet since previous work on perceptual narrowing has shown that Spanish 4.5-month-old infants were equally good at discriminating a Catalan speech contrast (/e-ɛ/) as their Catalan peers (Bosch & Sebastián-Gallés, 2003). We attribute the early representation of the vowel space to the mechanism that is behind phoneme acquisition. Distributional learning is based on accumulation of statistical information from the input. Our account on the early phonemic knowledge on vowels has highlighted the importance of distributional learning.

Another contribution of our experiment on vowel preference with infants is on the methodology we used. We adapted a behavioral paradigm that has been used to test infants' perception of phonotactics or stress patterns at 9 months (e.g. Jusczyk et al., 1993). In our experiment we adapted this paradigm into an eye-tracking setup for 4.5-month-olds. The advantages of such a method are that we eliminate any unintentional interference from the experimenter or the coders. In the classic headturn preference procedure, which is the technique typically used to investigate preferences, the experimenter has to perform online coding while the infant is being tested. This process entails that the experimenter must be fully concentrated on the looking behavior of the participants and leaves no room for error. Additionally, the videos of the participants have to be coded offline to measure the looking times accurately, which is a lengthy and tedious process. In our eye-tracking adaptation, the procedure was completely infant-controlled as it was contingent on the looking behavior of the participant. One disadvantage of the paradigm we used is that eye-trackers can be sensitive to movement. We hope that other researchers will benefit from our adaptation of the classic preference experiment setup and test their own hypotheses.

Although this was not the main goal of the thesis, we also addressed indirectly the importance of intonational information at this age. We tested 4.5-month-old infants on their capacity to discriminate Western Catalan from Spanish using natural (Experiment 2) and low-pass filtered (Experiment 4) sentences. Our hypothesis was that if infants used vowel distribution, they would fail to discriminate them as they have similar vowel distributions. Unexpectedly, we observed significant discrimination in Experiment 2 and marginal discrimination in Experiment 4, though both effects were small based on the values we obtained from the Bayes Factor analyses. In contrast, the Bayes factors we obtained in the experiments using segmental information were moderate to strong (Quintana & Williams, 2018). Our results from Experiment 2 and Experiment 4 are inconclusive, as it could be that infants have some initial but limited capacities to use intonation for language discrimination at 4.5-months of age. Converging evidence comes from previous research on the use of intonational cues to mark boundaries. It was found that infants at 4 months process intonational cues holistically while at 6 months they could use them individually and they preferred the prominent one in the native language (Seidl, 2007; Seidl & Cristià, 2008). Likewise, Chong et al. (2018) did not find discrimination at 5 months of a pair of languages that differ in their intonation. The other possibility is that the intonation cues were not particularly salient. Overall, we cannot draw any firm conclusions on the issue of intonation as our results with low-pass filtered sentences are at best weak. Seemingly our results contribute partially to the weak role of intonation at this age. Further research is required.

A contentious topic in psycholinguistics is that of the bilingual advantage in language processing. The basis of such an advantage is that bilinguals early on must separate their languages to master them. Previous research has not found any differences in the ability of monolingual and bilingual infants to discriminate between languages. Equal discrimination capacities have been found testing newborns with English and Tagalog sentences (Byers-Heinlein et al., 2010), 3.5-month-old infants with Basque and Spanish sentences (Molnar et al., 2014) and 4.5-month-old infants with Catalan and Spanish sentences (Bosch & Sebastian-Galles, 2001). It is noteworthy that although both populations can discriminate between languages, different processes might be involved. This has been indicated by electrophysiological data showing different patterns of activity between monolingual and bilingual infants presented with sentences in their native language and two unfamiliar languages (Italian, German) (Nacar Garcia et al., 2018). The findings from the present dissertation converge with previous results on language discrimination, i.e. we found that the capacity to discriminate Catalan and Spanish did not require bilingual expertise, suggesting that it might not be as hard to discriminate them (for a detailed description of bilingual language acquisition see Höhle et al., 2020; Sebastian-Galles & Santolin, 2020).

Due to the pandemic, we were forced to change our original research plan. We adapted to the present circumstances by conducting two experiments online with adults. Despite conducting these experiments ad hoc, some results are interesting and may deserve future research.

We investigated the role of intonational cues regarding language discrimination with adults. We found converging evidence that adults can rely on intonational cues to discriminate Western Catalan from Spanish, when they had access only to prosodic information. We also replicated findings that being a native speaker of a language is facilitatory in discriminating languages. The main contribution of the research reported in Chapter 3 was that we provide empirical evidence that two languages can be discriminated based on intonational cues. The intonational differences of the two languages have been mainly studied from a theoretical and an acoustic standpoint (Frota et al., 2007; Gibson, 2010; Ohannesian, 2005; Prieto et al., 2012; Wheeler, 2016). Previous studies have shown that adults rely more on segmental cues than prosodic ones; we found that they can also use prosodic cues alone when segmental information is reduced (Leemann et al., 2018; Van Bezooijen & Gooskens, 1999).

We also explored whether adults would be able to use global information (vowel distribution) or local information (phonotactics) when they were presented with lists of nonwords resembling their native distribution of vowels or not. Previous research has found that phonotactic information is pivotal in speech processing for adults. It has been shown that adults encode phonotactic permissibility implicitly when presented with speech, although they were not asked to focus on the combinations of sounds (Rossi et al., 2011; Toro et al., 2011). We tested two groups of participants with different language backgrounds, Catalan- and Spanish-dominant adults. Contrary to infants, we found an asymmetry between the two groups of participants suggesting that adults paid attention to local cues to judge whether a list was native-like or not. Participants had two types of information available, the vowel distribution (global cues) and the phonotactic patterns (local cues). The first type of information was informative for both groups of participants, while the second only for Catalan participants. Thus, we provide evidence that adults are insensitive to global distributional cues on vowels and are susceptible to local, phonotactic information. One critical observation is that even in the face of no useful phonotactic cue, Spanish adults were unable to disengage their focus from local to other cues that might have been more informative, i.e., vowel distribution.

6.3. Limitations and future research

6.3.1. Experimental caveats

Although the advantages and caveats for each experiment have been discussed throughout the previous chapters, some limitations presented here deserve further discussion. One general drawback is the use of low-pass filtered speech. This technique has been used widely in both infant and adult research. However, one of its shortcomings is that it does not reduce all segmental information equally. Some vowel information might be still left in the signal related to the first and the second formants. Low-pass filtering might also affect pitch if fundamental frequency goes over the cut-off threshold that has been set. In our language discrimination experiment using Eastern and Western Catalan, we found a clear non-discrimination in the low-pass filtered experiment. The p-value and Bayes factor were strongly in favor of the null hypothesis, suggesting that even if some information was still left in the signal, it was not adequate for infants to succeed in the task. Regarding the low-pass filtered discrimination experiment using intonational cues (Western Catalan, Spanish), we cannot reject the possibility that the statistically non-significant effect observed with the infants might be due to the stimuli. We recorded the stimuli with a female speaker who was instructed to produce the sentences in an infant-directed manner. It could be that by removing information above 400Hz, we also removed some pitch-related modulations. Though, this seems unlikely given that adults were able to discriminate the sentences above chance when they had access to the same intonational cues.

One caveat of both experiments conducted with adults was participants' selection. We were unable to control fully who participated in the experiment since they were conducted online. Concerning the experiment on language discrimination with adults (Chapter 3), the variability in the native language of the non-native participants impedes us from attesting whether intonation is perceived as general acoustic differences or whether participants can notice specific intonational cues that are used in their native language as well. Regardless of our efforts, a big percentage of our non-native participants had been exposed to the languages tested or even had some comprehension skills in Catalan and/or Spanish. We did not find a correlation between their discrimination scores and their self-reported comprehension scores, which indicated that nativeness and familiarity are not the same. Similarly, participant selection in Chapter 5 was not ideal given that most of the Catalan participants were bilingual in both languages tested, while most of the Spanish participants reported being fluent only in Spanish. This discrepancy is inherent to some degree to the linguistic background in Catalonia as most people are bilinguals. Although, our way of classifying participants' language dominance (based on the language of test) was not ideal, the big difference in performance between the two groups suggests it was reliable enough to observe a language-specific effect. An issue that we did not tackle in either adult experiment was to investigate the implication of being a speaker of both or only one of the languages tested. Future research can shed some light on the issue.

6.3.2. Future directions

To gain a better understanding of the phonemic knowledge infants have at their disposal, future research should explore a few issues that we did not get the chance to address here. Our results show that infants can discriminate sentences based on segmental information, but we do not know how fast infants can notice such differences. The maximum duration of one trial in our experiments was 28 seconds. Previous research designed to measure language recognition has shown that monolingual Catalan or Spanish infants can identify their language in 1173ms and 1344ms respectively when the sentences were natural (Bosch & Sebastian-Galles, 1997). Therefore, infants need a short amount of input to recognize a sentence as native or not. Future investigations could try to identify how many vowels infants need to listen to in order to know that a list of words or a sentence is native like.

An issue that we were unable to explore due to the pandemic is the nature of phonemic knowledge in bilingual infants. Our prediction was that bilingual infants would show no preference for one type of list over the other. This could be due to a common representation of the phonetic system at the age of testing or because they would consider both languages as native showing no preference for one over the other. Previous research on 4.5-month-old Catalan-Spanish bilingual infants showed that they attended equally fast to sentences in their native languages, i.e. showed no preference. The effect observed was the same regardless of language dominance. Though, the dominant languages were chosen based on the language that the mother used with the infant and no additional information was disclosed regarding their exposure to the two languages (Bosch & Sebastian-Galles, 1997). Another possibility is that infants' preference would be modulated by the degree of exposure to each language. To disentangle these possibilities, it would be fruitful to test bilinguals of the same age that have different degrees of exposure to the two languages. Balanced bilinguals (~ 50-50%) might not show a preference yet as they might need additional exposure to the two languages. Instead, Catalan or Spanish dominant bilinguals might show a preference more similar to their monolingual peers. Sebastián-Gallés & Bosch (2002) used this approach to study phonotactic knowledge at 10 months modulated by language exposure. If we observe differences depending on the degree of exposure, it could suggest that bilingual infants do not have a common representation, instead they need more input in order to have a clearer representation of the vowel system in the two languages.

Overall, we have provided the very first evidence that infants possess phonemic information before the first signs of perceptual narrowing. Future research should replicate our findings and extend them by using other combinations of languages to prove its generalizability. A topic that is worth exploring further is the type of information that infants have at 4.5 month of age and how it is organized.

An additional line of research that deserves further investigation is that of language discrimination with adults. It could be productive to test an additional group of foreign participants that has no contact with the testing languages, i.e. completely naïve participants. By doing so, one could claim with certainty that any effects observed are not due to familiarity. It could also be beneficial to run the same experiment but using the low-pass filtered sentences in Eastern and Western Catalan from Chapter 2. These sentences should not be discriminated since they do not contain any prosodic differences. If we do observe this pattern of results, it would support previous theoretical findings on the similarity of the two dialects but most importantly it would also support our findings with infants, who discriminated between the two in the natural speech condition only. Such study is under way in our laboratory.

Lastly, our experiment on the (in)sensitivity of adults to vowel distribution revealed an interesting asymmetry between Catalan and Spanish participants. We justified such asymmetry due to some tokens being phonotactically illegal in Catalan from the Spanish lists, which allowed them to detect lists that are native like successfully. In order to provide additional support that the reason behind such differences was the use of phonotactics, it could be fruitful to adapt this paradigm to record both behavioral and neuroimaging data in a laboratory setting. Participants can be recorded with EEG while performing the behavioral task. Previous literature has shown that adults perceive legal and illegal nonwords differently in a passive listening task. Particularly, they showed a stronger negativity effect for the legal structures. By collecting both behavioral and neuroimaging data, we could have additional evidence showing that Catalan participants have encoded phonotactic permissibility. Participants' performance could be analyzed in relation to the amplitude of the N400. One question that remains open is what would be observed in the Spanish group. Our hypothesis is that they would show similar N400 for all nonwords.

6.4. Conclusions

In the present dissertation we explored infants' phonetic knowledge at 4.5 months of age, a transitional time from general to language-specific speech perception. Our main contribution of the present dissertation is that we have provided evidence that infants have a representation of the vowel distribution of their native language before perceptual narrowing for vowels is evident. This knowledge allows them to discriminate between languages that differ in their incidence of vowels and to show a preference towards stimuli that resemble their native language. The possibility that infants have phonetic knowledge before the age of 6 months had been rejected repeatedly up until now. Thus, our findings enrich our understanding on the developmental trajectory of phoneme acquisition. It is the first step in filling the gap in the literature of language acquisition between neonates' and 6-month-old infants' capacities to perceive speech. Our findings on the sensitivity of infants to intonational cues are not conclusive. We also found that adults can use intonational differences to discriminate between languages, and although being native is not necessary for them to succeed, it is facilitatory. Lastly, adults are insensitive to vowel distribution information, at least at the word list level. The results from the present thesis provide pivotal information on infants' early perceptual abilities during the first months of life. Such knowledge is crucial to fully grasp the phenomenon of language acquisition.

7. Bibliography[20]

Abercrombie, D. (1967). *Elements of general phonetics*. University Press. https://books.google.es/books?id=bmQhAQAAMAAJ

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 72–77.

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078. https://doi.org/10.1121/1.4982246

Alcina, F. J., & Blecua, J. M. (1975). Gramática española. *Barcelona: Ariel, 1991*.

Allen, V. F. (1971). Teaching Intonation, from Theory to Practice. *TESOL Quarterly*, *5*(1), 73–81.

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical Basis for Speech Sound Discrimination. *Language and Speech*, *46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Aslin, R., Jusczyk, P., & Pisoni, D. (1998). *Speech and auditory processing during infancy: Constraints on and precursors to language.*

Barkat, M., Ohala, J., & Pellegrino, F. (1999). Prosody as a distinctive feature for the discrimination of arabic dialects. *6th European Conference on Speech Communication and Technology (EUROSPEECH'99).*

Bolinger, D. (1982). Intonation and Its Parts. *Language*, *58*(3), 505–533.

Bosch, L, & Sebastian-Galles, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual

[20] This list includes the work cited in the Introduction (Chapter 1) and Discussion (Chapter 6).

and bilingual environments. *Cognition*, *65*(1), 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1

Bosch, L, & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Bosseler, A. N., Taulu, S., Pihko, E., Mäkelä, J. P., Imada, T., Ahonen, A., & Kuhl, P. K. (2013). Theta brain rhythms index perceptual narrowing in infant speech perception. *Frontiers in Psychology*, *4*(October), 1–12. https://doi.org/10.3389/fpsyg.2013.00690

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–598. https://doi.org/10.1111/desc.12242

Bush, C. N. (1967). Some Acoustic Parameters of Speech and Their Relationships to the Perception of Dialect Differences. *TESOL Quarterly*, *1*(3), 20. https://doi.org/10.2307/3586196

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' Discrimination of Familiar and Unfamiliar Accents in Speech. *Infancy*, *16*(4), 392–417. https://doi.org/10.1111/j.1532-7078.2010.00050.x

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343–348. https://doi.org/10.1177/0956797609360758

Carbajal, M. J. (2018). *Separation and acquisition of two languages in early childhood: A multidisciplinary approach*.

Carbajal, M. J., Fér, R., & Dupoux, E. (2016). Modeling language discrimination in infants using i-vector representations. *CogSci*.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. https://doi.org/10.1016/0378-5955(94)00164-L

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-

specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561
Chong, A. J., Vicenik, C., & Sundara, M. (2018). Intonation Plays a Role in Language Discrimination by Infants. *Infancy*, *23*(6), 795–819. https://doi.org/10.1111/infa.12257
Christophe, A, & Morton, J. (1998). Is Dutch native English? Linguistic analysis by two-month-olds. *Developmental Science*, *1*, 215–219.
Christophe, A., Nespor, M., Guasti, M. T., & Ooyen, B. Van. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter Anne. *Developmental Science*, *2*(6), 211–220.
Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-directed Speech in the First Month after Birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x
Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Linguistics and Language Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015
Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., & Dupoux, E. (2014). Neural correlates of infant accent discrimination: An fNIRS study. *Developmental Science*, *17*(4), 628–635. https://doi.org/10.1111/desc.12160
Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, *12*(2), 58–71. https://doi.org/doi:10.1017/S0025100300002462
Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. In *Journal of Phonetics* (Vol. 11, Issue 1, pp. 51–62). Elsevier Science.
Dehaene-Lambertz, G., & Houston, D. (1998). Faster Orientation Latencies Toward Native Language in Two-Month-Old Infants. *Language and Speech*, *41*(1), 21–43. https://doi.org/10.1177/002383099804100102
Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303
Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *The Journal of Neuroscience*, *39*(29), 1828–18.

https://doi.org/10.1523/jneurosci.1828-18.2019
Fernald, A., & Kuhl, P. (1987). Acoustic Determinants of Infant Preference for Motherese Speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/https://doi.org/10.1016/0163-6383(87)90017-8
Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287–295. https://doi.org/10.3758/BF03205263
Frota, S., Imperio, M., Elordieta, G., Prieto, P., Frota, S., Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. (2007). The phonetics and phonology of intonational phrasing in Romance. In *Segmental and prosodic issues in Romance Phonology* (pp. 131–154). John Benjamins. hal-00142940
Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus, 13*(2), 247–275. https://doi.org/10.1515/prbs.2001.005
Frota, S., Vigário, M., & Martins, F. (2002). Language discrimination and rhythm classes: Evidence from Portuguese. *Proceedings of Speech Prosody 2002*, *April*, 319–322. http://sprosig.isle.illinois.edu/sp2002/papers.htm
Gasparini, L., Langus, A., Tsuji, S., & Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis. *Cognition*, *213*(May), 104757. https://doi.org/10.1016/j.cognition.2021.104757
Gibson, M. (2010). *Fonología del español: enfoque desde la teoría de optimidad. Issues in Spanish Phonology from an Optimality-Theoretic Approach*. Universidad Complutense de Madrid.
Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345
Grabe, E., & Low, E. L. (2002). *Durational Variability in Speech and the Rhythm Class Hypothesis*.
Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Grice, M., & Baumann, S. (2007). *An introduction to intonation-functions and models*. *18*(2), 73–83.

Grover, C., Jamieson, Do. G., & Dobrovolsky, M. B. (1987). Intonation in English , French and German : Perception and Production. *Language and Speech*, *30*(3), 277–295.

Hirst, D., & Di Cristo, A. (1998). A survey of intonation systems. In *Intonation systems: A survey of twenty languages* (Issue January 1998, pp. 1–44). Cambridge University Press. http://books.google.com/books?hl=en&lr=&id=LClvNiI4k0sC&oi=fnd&pg=PA1&dq=A+survey+of+intonation+systems&ots=fDM-zEhyHd&sig=EMpnXv8MCERe8-WA-f38XMpkYAA

Höhle, B., Bijeljac-Babic, R., & Nazzi, T. (2020). Variability and stability in early language acquisition: Comparing monolingual and bilingual infants’ speech perception and word recognition. *Bilingualism*, *23*(1), 56–71. https://doi.org/10.1017/S1366728919000348

Jara, C., Loccoz, C. M., & Peña, M. (2021). Infants exploit vowels to label objects and actions from continuous audiovisual stimuli. *Scientific Reports*, *0123456789*, 1–11. https://doi.org/10.1038/s41598-021-90326-z

Jun, S.A. (2010). Prosodic Typology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press: Oxford. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P.W., Cutler, A., & Redanz, N. J. (1993). Infants’ Preference for the Predominant Stress Patterns of English Words. *Child Development*, *64*(3), 675–687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants’ Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, 630–645.

Kitamura, C., Panneton, R., & Best, C. T. (2013). The development of language constancy: Attention to native versus nonnative accents. *Child Development*, *84*(5), 1686–1700. https://doi.org/10.1111/cdev.12068

Komatsu, M. (2007). Reviewing human language identification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4441 LNAI*, 208–228. https://doi.org/10.1007/978-3-540-74122-0_17

Komatsu, M., Arai, T., & Sugawara, T. (2004). Perceptual Discrimination of Prosodic Types and Their Preliminary Acoustic Analysis. *8th International Conference on Spoken Language Processing*.

Kuhl, P. K. (1979). *Speech perception in early infancy : Perceptual constancy for spectrally dissimilar vowel categories*. 1668–1679.

Kuhl, P. K. (1993). Innate Predispositions and the Effects of Experience in Speech Perception: The Native Language Magnet Theory. *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 259–274. https://doi.org/10.1007/978-94-015-8234-6_22

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, R. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5), 606–608.

Kuo, L.-J., Kim, T.-J., Yang, X., Li, H., Liu, Y., Wang, H., Hyun Park, J., & Li, Y. (2015). Acquisition of Chinese characters: the effects of character properties and individual differences among second language learners. *Frontiers in Psychology*, *6*(August), 1–10. https://doi.org/10.3389/fpsyg.2015.00986

Ladd, D. R. (2008). *Intonational_Phonology.pdf* (C. U. Press (ed.); June 2012). https://doi.org/https://doi.org/10.1017/CBO9780511808814

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich : New York. https://books.google.es/books?id=w_5YAAAAMAAJ

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews*, *81*, 158–166. https://doi.org/https://doi.org/10.1016/j.neubiorev.2016.12.012

Leemann, A., Kolly, M. J., Nolan, F., & Li, Y. (2018). The role of segments and prosody in the identification of a speaker's dialect. *Journal of Phonetics*, *68*, 69–84.

https://doi.org/10.1016/j.wocn.2018.02.001
Levis, J. M. (1999). Intonation in Theory and Practice, Revisited. *TESOL Quarterly*, *33*(1), 37. https://doi.org/10.2307/3588190
Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin and Review*, *18*(6), 1035–1041. https://doi.org/10.3758/s13423-011-0163-0
Liu, L., & Kager, R. (2017). Statistical learning of speech sounds is most robust during the period of perceptual attunement. *Journal of Experimental Child Psychology*, *164*, 192–208. https://doi.org/10.1016/j.jecp.2017.05.013
Lloyd James, A. (1940). *Speech signals in telephony,*. Sir I. Pitman & Sons.
Maddieson, I. (2013). Consonant-Vowel Ratio. In M. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. http://wals.info/chapter/3
Maidment, J. A. (1983). Language recognition and prosody: further evidence. *Speech, Hearing and Language: Work in Progress*, *1*, 133–141.
Maye, J., & Gerken, L. (2001). Learning Phonemes : How Far Can the Input Take Us ? *25th Annual Conference of the BUCLD, May 2011*, 480–490.
Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-11. https://doi.org/10.1016/s0010-0277(01)00157-3
McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x
Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 101–116.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

https://doi.org/10.1016/0010-0277(88)90035-2
Menard, L., Ouellon, C., & Dolbec, J. (1999). Prosodic markers of regional group membership: the case of the French of Quebec versus France. In J. Ohala (Ed.), *proceedings of the xivth international congress of phonetic sciences (icphs 99')* (pp. 1601–1604).
Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., & Sereno, J. (2017). Fetal rhythm-based language discrimination: A biomagnetometry study. *NeuroReport*, *28*(10), 561–564. https://doi.org/10.1016/j.physbeh.2017.03.040
Moftah, A., & Roach, P. (1988). Language recognition from distorted speech : Comparison of techniques. *Journal of the International Phonetic Association*, *18*(1988), 50–52. https://doi.org/10.1017/S0025100300003558
Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm Class Native Language Discrimination Abilities of Basque-Spanish Monolingual and Bilingual Infants at 3.5 Months of Age. *Infancy*, *19*(3), 326–337. https://doi.org/10.1111/infa.12041
Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U
Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica (Oslo, Norway : 1992)*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098
Muthusamy, Y. K., Barnard, E., & Cole, R. A. (1994). Reviewing Automatic Language Identification. *IEEE Signal Processing Magazine*, *11*(4), 33–41. https://doi.org/10.1109/79.317925
Nacar Garcia, L., Guerrero-Mosquera, C., Colomer, M., & Sebastian-Galles, N. (2018). Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. *Scientific Reports*, *8*(1), 2770. https://doi.org/10.1038/s41598-018-20824-0
Nazzi, T, Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology.Human Perception and Performance*, *24*(3), 756–766.
Nazzi, T, Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of

Rhythm and Familiarity. *Journal of Memory and Language*, *43*, 1–19.

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, *126*(3), 1440–1446. https://doi.org/10.1121/1.3158931

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-Timed vs . Syllable-Timed Languages. In *The Blackwell Companion to Phonology* (pp. 1–13). https://doi.org/10.1002/9781444335262.wbctp0048

Nolan, F. (2020). Intonation. In B. Arts, A. McMahon, & L. Hinrichs (Eds.), *The Handbook of English Linguistics, Second Edition* (Second Edi, pp. 385–405).

Nolan, F., & Jeon, H. S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658). https://doi.org/10.1098/rstb.2013.0396

Ohala, J. J., & Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody. *Problèmes de Prosodie: Expérimentations, Modèles et Fonctions*, *2*, 123.

Ohannesian, M. (2005). *La asignación del acento en castellano*. Universitat Autònoma de Barcelona,.

Ortiz-Mantilla, S., Hämäläinen, J. A., Musacchia, G., & Benasich, A. A. (2013). Enhancement of gamma oscillations indicates preferential processing of native over foreign phonemic contrasts in infants. *Journal of Neuroscience*, *33*(48), 18746–18754. https://doi.org/10.1523/JNEUROSCI.3260-13.2013

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences*, *107*(8), 3823–3828. https://doi.org/10.1073/pnas.0914326107

Peña, Marcela, & Melloni, L. (2012). Brain Oscillations during Spoken Sentence Processing. *Journal of Cognitive Neuroscience*, *24*(5), 1149–1164. https://doi.org/10.1162/jocn_a_00144

Peña, Marcela, Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience*, *32*(33), 11159–11163. https://doi.org/10.1523/JNEUROSCI.6516-11.2012

Pike, K. L. (1945). *The Intonation of American English*. http://search.ebscohost.com/login.aspx?direct=true&db=&AN=&site=eds-live

Polka, L., & Bohn, O.S. (1996). *A cross-language comparison of vowel perception in English-learning and German-learning infants*. *100*(1), 577–592. https://doi.org/10.1121/1.415884.

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., Molnar, M., Zhao, T. C., & Masapollo, M. (2021). Neurophysiological Correlates of Asymmetries in Vowel Perception: An English-French Cross-Linguistic Event-Related Potential Study. *Frontiers in Human Neuroscience*, *15*(June), 1–21. https://doi.org/10.3389/fnhum.2021.607148

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The Interplay Between Input and Initial Biases: Asymmetries in Vowel Perception During the First Year of Life. *Child Development*, *83*(3), 965–976. https://doi.org/10.1111/j.1467-8624.2012.01740.x

Prieto, P., Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*(6), 681–702. https://doi.org/10.1016/j.specom.2011.12.001

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *28*(3), 191–212. https://doi.org/10.1016/0028-2243(88)90030-5

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 1–8. https://doi.org/10.1186/s12888-018-1761-4

Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans XXIII*, *2*, 473–496.

Ramus, F, Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *15th ICPhs Barcelona*, *May 2014*, 1–6. http://cogprints.org/3079/

Ramus, F, Hauser, M. D., Miller, C., Morris, D., & Mehler, J.

(2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349–351.
Ramus, F, & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512–521. http://www.ncbi.nlm.nih.gov/pubmed/9921674
Ramus, F, Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.
Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. In *Annu. Rev. Lang. Acquis* (Vol. 2). https://doi.org/10.1075/arla.2.05ram
Reh, R. K., Hensch, T. K., & Werker, J. F. (2021). Distributional learning of speech sound categories is gated by sensitive periods. *Cognition*, *March*, 104653. https://doi.org/10.1016/j.cognition.2021.104653
Richardson, J. A. (1973). *The Identification by Voice of Speakers Belonging to Two Ethnic Groups*. The Ohio State University.
Roach, P. (1982). *On the distinction between ' stress-timed ' and ' syllable-timed ' languages*.
Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience*, *23*(7), 1752–1764. https://doi.org/10.1162/jocn.2010.21547
Schachner, A., & Hannon, E. E. (2011). Infant-Directed Speech Drives Social Preferences in 5-Month-Old Infants. *Developmental Psychology*, *47*(1), 19–25. https://doi.org/10.1037/a0020740
Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *PNAS*, *118*(7), 2001844118. https://doi.org/10.1073/pnas.2001844118/-/DCSupplemental.y
Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 28, Issue 4, pp. 974–989). American Psychological Association. https://doi.org/10.1037/0096-1523.28.4.974
Sebastian-Galles, N., & Santolin, C. (2020). Bilingual Acquisition:

The Early Steps. *Annual Review of Developmental Psychology*, *2*(1), 47–68. https://doi.org/10.1146/annurev-devpsych-013119-023724
Seidl, A. (2007). Infants’ use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48. https://doi.org/10.1016/j.jml.2006.10.004
Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x
Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639. https://doi.org/10.1121/1.4807565
Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: Further tests with a non-human primate. *Developmental Science*, *8*(1), 26–35. https://doi.org/10.1111/j.1467-7687.2005.00390.x
Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, *64*(2), 171–180. https://doi.org/10.1016/j.jml.2010.11.005
Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2005). Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 95–100. https://doi.org/10.1037/0097-7403.31.1.95
Trehub, S. E. (1973). Infants’ sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91–96. https://doi.org/10.1037/h0034999
Trehub, S. E. (1976). *The Discrimination of Foreign Speech Contrasts by Infants and Adults*. *47*(2), 466–472.
Tsuji, S., & Cristia, A. (2013). *Perceptual Attunement in Vowels : A Meta-Analysis*. https://doi.org/10.1002/dev.21179
Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, *18*(1), 31–48. https://doi.org/10.1177/0261927X99018001003

Vicenik, C., & Sundara, M. (2013). The role of intonation in language and dialect discrimination by adults. In *Journal of Phonetics* (Vol. 41). https://doi.org/10.1016/j.wocn.2013.03.003

Wagner, P. S., & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research. *ISCA Archive, February 2016*, 1–4.

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*(FEB). https://doi.org/10.3389/fpsyg.2014.00077

Werker, J F, & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Werker, J F, & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In *The Oxford handbook of developmental psychology* (pp. 909–925).

Werker, Janet F, V Gilbert, J. H., & Humphrey, K. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development*, *52*(1), 349–355.

Wheeler, M. (2016). *Catalan stress is iambic*. *February*.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. In *Journal of Phonetics* (Vol. 35, Issue 4, pp. 501–522). Elsevier Science. https://doi.org/10.1016/j.wocn.2007.02.003

Yeung, H. H., Chen, K. H., & Werker, J. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*, 123–139.

Yoshida, K. A., Maye, J., & Werker, J. F. (2010). *Distributional Phonetic Learning at 10 Months of Age*. *15*(4), 420–433. https://doi.org/10.1111/j.1532-7078.2009.00024.x