Tracking Phonetic-Learning Abilities across the Lifespan: Electrophysiological and Behavioural Perspectives.

Begoña Antonia Pericàs Herrero

A thesis submitted to University College London

in fulfilment of the requirements for the degree of

Doctor of Philosophy.

I, Begoña Antonia Pericàs Herrero, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Begoña Antonia Pericàs Herrero.

Abstract

This thesis presents three experiments aimed at developing time-effective techniques for studying phonetic/phonological learning across the lifespan, using both behavioural and electroencephalography (EEG) measures. The first experiment investigates phonetic learning of the standard Southern British English (SSBE) vowel set in Spanish-Catalan children (6-12 years) using a computerised version of the *Memory/Concentration* card game. The first study was conducted in Majorcan schools and included three training sessions, with pre- and post-training tests of vowel discrimination using a three-alternative forced-choice oddity task. Results revealed significant improvements in SSBE vowel discrimination, with older children from bilingual home environments benefiting the most from the training program. This demonstrates that the *Memory* card game was an effective tool for child phonetic training, highlighting the influence of age and linguistic background on phonetic learning. The second experiment explores a novel EEG technique to map listeners' perceptual vowel spaces. Participants completed an EEG recording with an oddball paradigm designed to elicit Mismatch Negativity (MMN) responses, alongside behavioural vowel identification and category goodness tasks. Machine learning analysis of MMN responses showed potential for reducing testing time, since they gathered sufficient statistical power to demonstrate reliable MMNs after a single deviant per deviant type. However, the MMNs did not show clear phonological contributions. As a result, the oddball paradigm was modified and the third experiment uses this revised MMN approach to compare vowel perception amongst native British English speakers and native Spanish speakers who acquired English as a second language in adolescence or adulthood. The results demonstrated clear phonological contributions to the generated MMNs, however the observed MMN amplitudes did not mirror the native vowel repertoires of either speaker group. Together,

these experiments advance methodologies for testing phonetic training and cross-language vowel perception, facilitating the acquisition of larger datasets to enhance our knowledge on phonetic learning across the lifespan.

Impact Statement

The research outlined in the current thesis examines how learners perceive and acquire non-native speech sounds, considering both targeted phonetic training in children and cross-language vowel perception in adults. In particular, my work on child second language phonetic training highlights the benefits of early exposure to non-native sounds in the classroom. Research suggests that structured training, incorporating high-variability input, audiovisual cues, and interactive learning, can significantly enhance children's ability to distinguish and produce new phonemes. These findings have practical applications for language education, speech therapy, and curriculum design, supporting the development of evidence-based interventions that enhance phonetic learning outcomes across different stages of life. Moreover, understanding cross-language phoneme perception is crucial for improving language learning and teaching methods. My research aims to develop a tool to shed light on how learners process non-native speech sounds, revealing factors that influence their ability to distinguish and acquire them. These findings have implications for language education, helping educators design more effective perception training and curriculum adaptations. Additionally, this work contributes to our understanding of speech perception in multilingual contexts, benefiting fields such as linguistics, cognitive science, and speech therapy.

Acknowledgements

I'm incredibly grateful to my supervisor, Paul Iverson, for all his support throughout my PhD. Thanks for trusting me to run EEG studies and for helping me become confident and independent in the lab. I'm grateful for your flexibility during the times when testing there wasn't possible, and especially for adapting the memory card game code so I could test in schools. Thanks for being great fun at conferences and for giving me the chance to share my work in Spain and Canada, as well as for helping me get the research visit funding to go all the way to Japan.

I'd also like to thank Jo Taylor, Emma Holmes, and Carolyn McGettigan for their insightful and constructive feedback during my upgrade. A special thank you to Jo, my second supervisor, for her support in the early stages of data collection, for volunteering as a participant herself, and for giving me the opportunity to teach her students what I had learned about EEG. I'm also very grateful to Gordon Mills for all his help in the lab. His patience and expertise made a big difference, especially in the trickier moments.

All my lab mates, were a brilliant support network from the moment I joined the group. Thanks to Ana Campos and Magda Rojas for their constant help with countless tasks over these past four years. I'd also like to thank my SHaPS office colleagues, Ella Gregory, Eri Iwagami, Han Wang, Hannah Wilt, Rongru Chen, Ruohan Guo, and Ziyun Zhang, for always lending a hand when I needed it. And a special thanks to Jonas Huber, for being the best senior I could have asked for!

I am very thankful to Maria José Rivero and the teaching staff at CEIP Miquel Duran i Saurina and CEIP Llorenç Riber for kindly allowing me to carry out my training study in their schools and for all their support throughout the process. I had such a cute bunch of participants! Thank you to all the children who took part and did such an excellent job learning English vowels.

Thanks also to all the participants in my EEG studies, not only for showing up and staying still, but for not falling asleep during 45 minutes of listening to random vowels. True heroes of science!

I also want to thank my friends for being so supportive throughout this journey and for always being willing to help out. To my friend Cata Trama, thank you not only for becoming so skilled at post-EEG hair washing that you ended up as my default guinea pig, but also for always being there to help me rehearse my conference presentations. Carlota Quetglas, Dylan Springer, and Verónica Escobar, thank you as well for volunteering your brains for science. Catibel Mateu, Maria Batle, and Marga Batle, thank you for recording your Spanish and Catalan vowels for this project, and Adrià Daviu, Amelia Sirera, and Marie Claire Fischer, I'm very grateful for the long library sessions that helped bring this thesis to life. I also want to thank Andrei Savitski, Hannah Copeland, Clarissa Newman, Bhasvini Kul Kaur, Annabel Samuel, and Yasmin Feizal, for being such a fun and reliable source of distraction and laughter throughout the PhD.

To my uncles and aunt, Antonio Herrero, Juan de Amador, and Coral Herrero, thank you for testing my games and vowel tasks and for offering such helpful feedback.

And of course, to Mum and Dad, thank you for always being in my corner, for joining forces to cheer me up whenever London got too cold and rainy, and for being the best team I could have ever asked for. None of this would have been possible without you.

Table of Contents

List of Figures

List of Tables

1. General Introduction

Nowadays, many of the worlds' linguistic contexts expose the members of their community to multiple languages on a daily basis, increasing the number of bilingual or multilingual speakers worldwide. For instance, in communities like the island of Majorca, multiple languages are recognized for official and public use; while Spanish is the official language of the country, the Majorcan dialect of Catalan holds official status on the island. Bilingualism also frequently occurs at the individual level; for example, many families include members who have migrated to the family's country of residence, resulting in children learning one language at home and another in school. Exposure to multiple languages shapes listeners' speech processing abilities in various ways. Young infants exposed to multiple languages on a daily basis must learn to differentiate between their languages, leading to more refined language discrimination abilities compared to monolingual infants (e.g., Sebastián-Gallés et al., 2012). Some studies show that bilingual infants exhibit distinct patterns of perceptual attunement to their native language. That is, while monolingual infants typically lose the ability to discriminate non-native speech contrasts after six to eight months of age (Burns et al., 2007), bilingual infants retain the capacity to distinguish phonemic contrasts in both of their languages throughout their first year and beyond (Sundara et al., 2008; Sundara & Scutellaro, 2011). In childhood, immersion exposure to an L2 can lead to native-like perceptual abilities of L2 phonemes (McCarthy et al., 2014), and the acquisition of non-native phonetic knowledge is faster than in adulthood (Tsukada et al., 2005). Nevertheless, our understanding of monolingual language development remains significantly more comprehensive than that of multilingual language development.

The three experiments described in this thesis aimed to create time-effective techniques of testing native and non-native phoneme learning, and cross-language vowel perception across the lifespan. Initially, the current thesis aimed to explore early phonetic learning as, despite decades of research on this field, there is no clear understanding of the processes that enable infants to learn the sounds in speech. This is due to the nature of infant data as they are unable to sit through long experiments, and testing perception of several phonetic contrasts is often time-consuming. As a result, our knowledge on infant speech is based on studies of perception of few phonetic contrasts. The initial aim of this thesis was to create an electrophysiological (EEG) tool that would shorten the length of testing infant perception of multiple phonetic contrast. However, due to severe Covid-19 restrictions, which prevented laboratory-based infant testing, the scope of this thesis was adapted to include child and adult perceptual studies. The first study of the current thesis involves a phonetic training program delivered in a classroom setting, to over a hundred children. The second and third studies include two adult EEG studies aimed at evaluating a time-efficient analytical framework for the mismatch negativity (MMN) response, which was used to explore cross-language vowel perception in English and Spanish listeners. While the focus of this thesis remains on phoneme learning across the lifespan, it also considers learning in multilingual contexts, as most participants, excluding those in the second study, used more than one language in daily life.

Different aspects of language processing vary in difficulty for learners, and the learners' age affects how these challenges influence learning. L2 speech learning is generally very challenging, and late learners often speak their non-native language with non-native accents. In contrast, infants growing up bilingual can master the phonetics and phonology of both of their languages with relative ease. At this point, it important to define what we mean by the phonetics and phonology of language. A language's phonetic repertoire consists of the sounds used to produce words whereas its phonology governs how these sounds are organised and assigned meaning to form words. The current review will focus on phonetic learning, as it is of most relevance to the studies described in this thesis. The current chapter will provide a review of speech learning throughout the lifespan, focusing on how bi- or multilingualism affects speech perception abilities and the acquisition of the sound systems in language. It begins by explaining how phonetic perception is tested in the laboratory, and continues discussing research on infants raised in bilingual environments, emphasizing speech development during the pre-educational years. It then describes phonetic learning in childhood, drawing evidence from child studies of L2 phoneme learning. Lastly, it provides an overview of additional language learning and bilingual speech perception in adulthood, and concludes by providing a summary of the three studies presented in the next chapters. At the end of this review, it will be clear why further research is essential to gathering more data and constructing a complete narrative of phonetic learning development, as well as how this thesis contributes to that effort.

1.1. Tasks Measuring Phonetic Perception

The studies included in this thesis measure phonetic perception via behavioural and electrophysiological tasks. Accurate speech perception involves the recognition of phonetic features that serve to distinguish phonemes and allow listeners to make sense of words (Cutler, 2012). In children and adults, perception of phonetic segments or contrasts relies not only on recognising differences in the acoustic signals that allow for the categorisation of phonemes but also accessing said phonetic categories to identify the stimulus (C. Best & Tyler, 2007; Kuhl, 2004). Phoneme perception tasks prompt participants to reveal this internal perceptual process by producing a measurable response, and these paradigms can be described in terms of the tasks involved, the test structure, and the stimuli that are presented to the participants.

There are two general kinds of experimental tasks used in phonetic perception experiments: identification and discrimination tasks. In an identification task, participants are presented with recorded speech stimuli, one at a time and are prompted to classify each one into a phonetic category (Pisoni & Tash, 1974). This can be done either through an open-set format, where they provide a spoken or written response freely, or a closed-set format, where they choose from predefined options (such as orthographic symbols or key words) given by the experimenter. In a discrimination task, participants hear two or more stimuli, and must indicate the relationship between them, typically deciding whether they are the same or different (Gottfried, 1984). Many variations of these tasks have been used in child and adult studies of native and non-native phoneme perception.

The current thesis uses a combination of identification and discrimination tasks. In Chapter 2, Spanish-Catalan children are trained on the southern British English vowel set using a computerised version of the Memory card game (also known as Concentration). In our version of the game, the cards are linked to minimal pair words and a match consists of two cards linked to the same word, spoken by two different talkers. This game provides a learning context similar to a category discrimination task (Gottfried, 1984), as children must ignore the acoustic differences that do not help them decide whether the words are the same or different and must focus on the features that distinguish phonetic categories in the English language. To measure phonetic learning from pre- to post-training, we used an Oddity discrimination task, which is a more traditional way of measuring phonetic perception. Discrimination tasks vary in complexity, with the simplest being the AX task, where two stimuli are presented sequentially, and listeners judge whether they are the same or different (Werker & Logan, 1985). While such tasks can effectively assess basic auditory sensitivity, they are susceptible to response bias and involve relatively low cognitive demands. In contrast, Oddity tasks present three stimuli per trial (i.e., two from the same category and one from a different one) and require listeners to identify the stimulus that differs. This structure introduces greater memory load and stimulus uncertainty, as the position of the odd item varies across trials. Because of this, the Oddity task taps not only into auditory discrimination but also into categorization and implicit phonetic labelling (Hansen Edwards

& Zampini, 2008). As such, it is well-suited for detecting changes in phonetic representations that may emerge as a result of training.

A different kind of discrimination task is included in the electrophysiological experiments in Chapters 3 and 4, namely to Oddball task. . In this paradigm, repeated exemplars of a phonetic category are used as the Standard sound. Occasionally, an exemplar of a different phonetic category appears, introducing a Deviant trial. The listener's role is to detect and indicate when they perceive a shift from the Standard category to the Deviant category. The Oddball task is comparable to the AX task in terms of memory demands and stimulus uncertainty, and it is believed to primarily assess auditory discrimination skills while minimizing the influence of categorization processes (Näätänen et al., 2007a). In Chapters 3 and 4 we presented our participants with an Oddball paradigm whilst they watched muted cartoons to record neural indices of their perception of the IPA vowel set. Additionally Chapters 3 and 4 include a behavioural vowel identification task to evaluate how the IPA vowels are consciously identified and categorised. This allowed us to compare the behavioural and neural data and draw correspondences between both sets of data.

The speech stimuli that we used in the experiments of this thesis included recordings of real words by native speakers and synthetic vowel stimuli. Choosing the right speech stimuli is very important and often depends on the empirical questions being addressed. In our phonetic training study in Chapter 2, we included real words spoken by native southern British English speakers because these stimuli have been shown to be most effective in training (Iverson & Evans, 2009a) and they would allow children to improve their perception of English vowels in the real world. When hearing real words, second language (L2) listeners, may rely on their existing lexical representations during the perception task (Flege, 1995). In contrast, distinguishing between nonsense items depends solely on phonetic and phonological knowledge, independent of the lexicon (So & Best, 2010). Using real words can blur the distinction between lexical and phonetic influences but since real-world speech perception naturally involves both lexical and phonetic processing, studies using real words provide a more realistic representation of the challenges L2 learners face in everyday communication (McClelland & Elman, 1986).

However, Chapters 3 and 4 included a very different stimulus set (i.e., isolated and synthesised IPA vowels). Cross-language studies often utilize computer-generated synthetic speech, which allows researchers to precisely control which acoustic features vary and which remain unchanged (Bradlow & Bent, 2002). Nevertheless, even high-quality synthetic speech tends to sound somewhat unnatural. In our experiments we used a set of 19 IPA vowels spoken by a female phonetician which were synthesized to maintain a constant pitch. In Chapter 3 we were interested in mapping out southern British English speakers' pre-attentive vowel spaces, and these synthesised IPA vowels allowed us to create a suitable Oddball paradigm. In Chapter 4, we looked at cross-language vowel perception between English and Spanish speakers using a similar paradigm and this controlled set of synthesised vowels allowed us to elicit MMN responses in our participants and capture their perceptual abilities at the pre-attentive level.

1.2. Phonetic Learning in the First Year of Life

Although this thesis does not include infant participants due to COVID-19 constraints, infant studies remain foundational in the field of phonetic learning and provide important context for interpreting results in older populations. In infant research, perception is typically inferred through physiological or behavioural responses. Two of the most commonly used tasks are the conditioned head-turn procedure and the habituation/dishabituation paradigm. In the conditioned head-turn procedure, infants sit on their caregivers' lap and are presented with a visual stimulus on one side of the room and a sound on the opposite side. Initially, the sounds are presented randomly; later, researchers condition the infant to turn their head toward the sound source when a specific sound changes. Consistent head turns following the change suggest the infant can discriminate between the sounds. In the habituation task, infants are repeatedly exposed to the same sound until their attention declines. A new sound is then introduced, and researchers measure whether the infant shows renewed interest, indicating detection of the change. These tasks highlight the challenges of measuring speech perception early in life and emphasise the need for time-efficient, developmentally appropriate methods; challenges that are central to the present thesis and addressed through the use of EEG tools.

Before infants speak their first words, they already possess foundational knowledge of their linguistic environment. Research shows that, even before birth, foetuses can perceive prosodic features such as rhythm, which helps newborns discriminate between languages from different rhythmic classes (Nazzi et al., 1998). While monolingual newborns show a preference for their native language (Moon et al., 1993), bilingual newborns exposed to two languages in utero show no clear preference, attending equally to both (Byers-Heinlein et al., 2010). Importantly, this sensitivity extends even to languages from the same rhythmic class, as shown in 4-month-old Spanish-Catalan infants who successfully differentiated between their two native languages (Bosch and Sebastián-Gallés, 2001). These findings suggest that early sensitivity to rhythm supports bilingual infants' ability to separate and attend to both of their languages from birth.

As infants develop an early grasp of rhythm, the amount and quality of input in each language become crucial for acquiring segmental properties of speech, a factor that remains significant in adult learners (e.g., Flege & Bohn, 2021). Bilingual infants typically receive less input in each language than monolinguals, which may impact the timing and development of phonological categories. This interacts with perceptual attunement (i.e., the reorganisation of the perceptual system during the first year of life). From birth, infants are sensitive to subtle differences in many phonetic contrasts, even those that extend to non-native contrasts that adult listeners have difficulty differentiating (Werker, 1989). These initial sensitivities change across the first year of life so that, at around 6 months of age, monolingual infants start to tune into their native language. At 6 months, they start showing a decline in discrimination of non-native phonetic contrasts (e.g., Werker & Tees, 1984) and an improvement in discrimination of native ones (e.g., Kuhl et al., 2006). This process of perceptual attunement or perceptual narrowing, is seen as the first step towards phoneme categorisation, a necessary milestone for learning words. Moreover, these early experiences shape the perceptual space in ways that affect speech learning throughout the lifespan (Iverson et al., 2003), including the non-native phoneme learning investigated in this thesis.

Research has shown that monolingual and bilingual infants exhibit different patterns of attunement, with distinct developmental trajectories observed for vowels versus consonants. The first investigation into consonant discrimination in bilingual infants compared English–French bilingual infants with English monolingual infants (Burns et al., 2007). English and French employ voicing as a distinctive phonetic feature, but the voice-onset-time (VOT) boundary distinguishing voiced and voiceless labial consonants differs between the two languages. In this study, three syllables with varying VOT values were tested: one with a short VOT (interpreted as voiced /ba/ by adult listeners in both languages), one with an intermediate VOT (perceived differently across languages, as /pa/ by French-speaking adults and /ba/ by English-speaking adults), and one with a long VOT (heard as aspirated /pha/ in both languages). Using a visual fixation procedure, infants were first habituated to the syllable with the intermediate VOT and then presented with the other two syllables during the testing phase. At 6–8 months of age, both monolingual and bilingual infants exhibited similar responses: they only showed renewed attention, or dishabituated, to the aspirated syllable with the long VOT. However, by 10–12 months and 14–20 months, differences emerged. While monolingual English infants continued to show the same pattern as the younger group, their bilingual peers dishabituated to both contrasts. These results suggest that bilingual and monolingual infants do not differ in their perception of these sounds at 6–8 months, supporting the idea that perceptual tuning for consonants develops during the latter half of the first year. By the end of the first year, however, clear distinctions between monolingual and bilingual infants became apparent. Monolingual English infants categorized the intermediate VOT sound as voiced, whereas bilingual infants demonstrated a more nuanced discrimination. Like monolingual English infants, they distinguished between the intermediate and long VOT sounds, but they also differentiated between the intermediate and short VOT sounds, a pattern consistent with that of French monolingual infants of the same age.

A subsequent study examined how French and English monolingual infants, as well as French–English bilingual infants, discriminated between two versions of a syllable starting with the coronal /d/ sound, produced by either a Canadian English or Canadian French speaker (Sundara et al., 2008). The /d/ sound differs slightly between English and French, with French /d/ being more dental and English /d/ more alveolar in its articulation. Previous research with adults had demonstrated that English monolingual and English–French bilingual adults were better at distinguishing these subtle articulatory differences compared to French monolingual adults (Sundara & Polka, 2008). In this study, infants aged 6–8 months and 10–12 months were habituated to stimuli from either a French or English speaker and then tested with stimuli from different speakers of both languages. At 6–8 months, all infants, regardless of language background, were able to discriminate between the French and English /d/ sounds. However, by 10–12 months, language background began to influence performance: while English monolingual and bilingual infants continued to discriminate the sounds, French monolingual infants no longer showed a discrimination response. These findings suggest that monolingual and bilingual infants follow similar developmental trajectories in sound discrimination, and that the ability to distinguish contrasts is maintained if at least one of the infant's languages supports it.

A series of studies investigated the neural mechanisms underlying the discrimination of VOT differences in English monolingual and Spanish–English bilingual infants, focusing on native and non-native sound contrasts using event-related potentials (ERP) and magnetoencephalography (MEG) (Ferjan Ramírez et al., 2017a; Garcia-Sierra et al., 2011; Rivera-Gaxiola et al., 2005). These studies employed three stimuli from a /d/–/t/ voicing continuum: a pre-voiced stimulus with a negative VOT (characteristic of Spanish voiced /d/), a stimulus with a long positive VOT (characteristic of English voiceless /ta/), and a stimulus with a short positive VOT (categorized as voiceless /ta/ in Spanish and voiced /da/ in English). A double-oddball paradigm was used, with the short positive VOT sound as the standard, and the neurophysiological mismatch response (MMR) to the two deviants was measured. The mismatch response is elicited when an infrequent stimulus is perceived as different from a recurring stimulus and is explained in detail in Chapter 3. For Spanish–English bilingual infants, no MMR was observed for either the English or Spanish deviants at 6–9 months, but a negative MMR emerged for both deviants by 10–12 months (Garcia-Sierra et al., 2011). Notably, the strength of the MMR was influenced by language exposure: infants with more English exposure showed a stronger response to the English contrast, while those with more Spanish exposure responded more strongly to the Spanish contrast. These findings are surprising, as other studies have shown that 6-month-old bilingual infants can discriminate native speech contrasts (Burns et al., 2007; Sundara et al., 2008). Moreover, even monolingual English infants did not show consistent neurophysiological responses to this English–Spanish voicing contrast (Rivera-Gaxiola et al., 2005).

A direct comparison between monolingual English and bilingual Spanish–English 11-month-olds using MEG revealed distinct patterns (Ferjan Ramírez et al., 2017).

Monolingual infants showed no difference in MMR for the English and Spanish contrasts in the early time window of analysis, but a larger MMR for the English contrast in a later window. In contrast, bilingual infants exhibited a stronger response to the English contrast in the early window and a stronger response to the Spanish contrast in the later window. Group comparisons indicated no differences in responses to the English contrast, but bilingual infants showed stronger responses to the Spanish contrast in both time windows compared to monolinguals. Meaning that while, monolingual infants demonstrated sensitivity to English, bilingual infants exhibited sensitivity to both of their languages. Moreover, the authors argued that the neural evidence suggests that this dual sensitivity in bilingual infants arises from a more gradual shift from acoustic to phonetic processing.

The reviewed evidence reveals a complex picture of phonetic development in bilingual infants as there are shared commonalities with monolingual trajectories (Burns et al., 2007; Sundara et al., 2008) yet also shows marked differences in neural responses to speech sounds (Ferjan Ramírez et al., 2017b; Garcia-Sierra et al., 2011). For instance, bilingual infants may experience an extended period of language-general consonant discrimination before shifting to language-specific patterns (Ferjan Ramírez et al., 2017). Moreover, studies using the MMR have shown that bilinguals' neural responses vary depending on language dominance and relative exposure with stronger MMRs being observed in the dominant language (Garcia-Sierra et al., 2011). These inconsistencies highlight the need for further research comparing native and non-native consonant discrimination in both monolingual and bilingual infants. In particular, there is a need for testing paradigms that can efficiently capture perception of multiple phonetic contrasts across different language backgrounds. Chapter 3 addresses this gap by evaluating a time-efficient mismatch negativity (MMN) paradigm that holds promise for application in infant and lifespan research.

In particular, the MMN paradigm described in Chapter 3, measures perception of the vowel space. While consonant perception studies often reveal a clear attunement pattern, with improved discrimination of native contrasts and reduced sensitivity to contrasts that are not used in the infants' languages, vowel perception studies show greater variability in results and few studies show the aforementioned attunement pattern (Tsuji & Cristia, 2014). The first study to explore vowel perception development in bilingual infants compared Spanish and Catalan monolingual and bilingual infants on their ability to discriminate between the vowels /ɛ/ and /e/, which are distinct phonological categories in Catalan but not in Spanish (Bosch & Sebastian-Galles, 2003). Using the head-turn preference procedure, infants were familiarized with one vowel and then tested with both. At 4 months, all groups (i.e., monolingual Spanish, monolingual Catalan, and bilingual infants) could discriminate the contrast. By 8 months, only monolingual Catalan infants continued to show discrimination, while at 12 months, bilingual infants regained the ability to distinguish the vowels. These results suggest that vowel perception at 4 months is not yet influenced by language exposure, and by 8 months, monolingual infants had attuned to their respective vowel systems. The bilingual infants, however, appeared to follow a U-shaped trajectory, only fully attuning to the Catalan vowel system by 12 months.

Challenging this interpretation, a later study used an anticipatory eye movement task to test Catalan–Spanish bilingual and Catalan monolingual 8-month-olds with the same vowel contrast (Albareda-Castellot et al., 2011). Both Catalan monolingual and bilingual infants improved their accuracy in anticipating the location of a visual target based on the vowel presented, while Spanish monolingual infants did not show similar improvement. Unlike the earlier findings, these results do not support a delay in perceptual attunement for bilingual infants. Instead, they suggest that bilingual infants' performance may be more sensitive to the specific experimental paradigm used, potentially indicating that their ability to discriminate between the tested sounds is less robust than that of monolinguals and more susceptible to variations in task demands. The same vowel contrast, which is phonemic in English, was tested in monolingual English and bilingual Spanish–English infants aged 4 and 8 months (Sundara & Scutellaro, 2011). Using a visual fixation paradigm, infants were habituated to one vowel and then tested with both vowels. Both monolingual and bilingual infants at 4 and 8 months successfully discriminated the contrast. Consistent with the findings of Albareda-Castellot et al. (2011), these results suggest that bilingual infants retain the ability to discriminate a vowel contrast, even when it is phonemically relevant in only one of their languages. The MMN tool designed for this thesis can be used to test vowel perception in passive listening experiments, reducing the task demands and allowing us to capture perceptual abilities even in young participants.

1.3. Phonetic Development in Childhood

Bilingual phonetic development in childhood has received far less attention than bilingual speech learning in infancy and it is the central topic of Chapter 2. To link infant and child speech development, Werker and Curtin (2005) proposed the PRIMIR (Processing Rich Information from Multidimensional Interactive Representations) model of speech perception development and word learning.. PRIMIR proposes three interactive representational spaces of speech: the General Perceptual Space, which holds detailed phonetic and indexical information; the Word Form Space, where recurring sound sequences are stored; and the Phoneme Space, which emerges as vocabulary grows and learners begin to abstract stable phonemic categories. Over time, this system supports increasingly fine-grained phonemic representations, which are crucial for accurate word recognition and production. Rather than developing linearly, these representational spaces interact dynamically, adapting to ongoing linguistic experience and guiding the refinement of phonetic and phonological knowledge throughout childhood. That is, once children have acquired enhanced sensitivities to the phonemes in their native languages, they learn to categorise these sounds into phonological categories, a process that may refine until adolescence (Hazan & Barrett, 2000).

Some cross-sectional studies that have applied the same experimental approach across different age groups reveal a clear developmental trajectory of phonetic categorisation. For instance, Sundara et al., (2008) examined how monolingual English speakers at different ages (i.e.,10- to 12-month-old infants, 4-year-old children, and adults) discriminated the English /d/-/ð/ contrast. They used a modified version of the conditioned head-turn procedure, where older children and adults pressed a button when detecting a change in the stimulus. Results showed that 4-year-olds outperformed infants in distinguishing the contrast but were not as consistent as adults. Similar patterns of development have been observed for non-native phonemic contrasts, using the same modified head-turn procedure. Werker and Tees, (1983) examined the perception of native and non-native phonemes in English-speaking 4, 8 and 12-year olds. In particular, they tested perception of two Hindi contrasts: the dental-retroflex stops / t̪a/-/ ʈa/ and the voiced-voiceless aspirated dental stop /dʰ/-/tʰ/, as well as the English /ba/-/da/ contrast. Results demonstrated an overall worse discrimination performance for non-native contrasts compared to native ones and, in the voicing contrast, the 4 year-olds were the worst performers, indicating that in the younger years of childhood, non-native phoneme perception is most difficult.

To further examine children's phonetic categorization patterns, other studies have used identification paradigms, which prompt children to not only perceive differences between phonemes but also attach labels to them. This research demonstrates that children and adults use different acoustic cues when categorizing phonemes, a concept known as 'perceptual cue weighting' (Nittrouer, 1996; Nittrouer & Miller, 1997). Nittrouer's Developmental Shift Weighting hypothesis (DSW, 1996) suggests that children initially focus more on broad measures like formant transitions, which are more noticeable than static cues such as duration. Studies supporting this hypothesis show that children are more responsive to formant transition cues in voicing and vowel identification (Ohde & Haley, 1997; Sussman & Carney, 1989). Nittrouer (1996) also proposes that as children's vocabularies expand, they require a more detailed representation of speech sounds, leading to a shift from general to specific adult-like perception patterns. Similarly, Burnham's robust-fragile hypothesis (1986) states that speech perception development is influenced by the prominence of acoustic contrasts. Fragile contrasts refer to sound differences that are both acoustically subtle and uncommon in the world's languages. If a child's native language does not include these distinctions, the ability to tell them apart is expected to diminish within the first year of life. In contrast, robust contrasts are characterized by clear and easily noticeable acoustic differences. They appear frequently across languages, and tend to be reliably distinguished by children up to school age, even without direct exposure.

Researchers have also investigated non-native phonetic development in a target language that children are learning. Studies on sequential bilingual children, that is, children who acquired their L2 after they had started learning their native language (L1), have focused on comparing children's and adults' ability to discriminate non-native phonemes. There are two main differences between children and adults, which influence L2 learning. Children have greater neuroplasticity, which is beneficial for learning, and unlike adults, they do not have a fully built L1 phonological system, which could interfere with the phonology of the L2 (Werker & Tees, 2005). In the real world, children are more likely to achieve a native-like accent than adults (Flege et al., 2003) and the rate of L2 phonetic learning is generally faster. Tsukada et al., (2005) compared how well Korean adults and children aged 9 to 17 could perceive English speech sounds after living in America for either three or five years. After three years, children were already significantly better at distinguishing English sounds than adults and their ability to discriminate English phonemes improved after 5 years of living in America. Therefore, although laboratory studies demonstrate that young children find it harder than adults to perceive non-native phonemes they are not learning, second language phonetic learning is ultimately faster and easier in childhood.

To examine the role of formal education in sequential bilinguals' non-native phonemic learning abilities, McCarthy et al., (2014) tested four-year-old Sylheti-English children who spoke Sylheti at home and were schooled in English. Their results showed that, before starting kindergarten, Sylheti-English children were less skilled than their monolingual peers at distinguishing between voiced and voiceless English plosives. However, after one year of kindergarten, their abilities improved significantly, reaching the same level as the monolingual group. The children in this study may have benefitted from being of a younger age than those in Tsukada et al., (2005) when learning English phonemes. Nevertheless, a study with 11-year-old Turkish-German sequential bilinguals who had been immersed in a German environment before they turned 4 years old, showed that their perception of German vowels was significantly poorer than that of German monolingual children (Darcy & Krüger, 2012). However, unlike the children in McCarthy et al., (2014), who attended a monolingual kindergarten, the Turkish-German children were schooled in a Turkish-German bilingual programme. Therefore, they received less input in their L2 than did the Sylheti-English children.

These findings highlight the importance of quantity of L2 input in child phonetic learning, demonstrating that near-native or even, native-like levels of phonetic perception are only achieved with substantial amounts of L2 input, even when the L2 is acquired from the age of 4 years. Unsurprisingly, Goriot et al., (2020) showed that English immersion education, starting from 4 years of age may not be sufficient for native-like phonetic perception attainment. In their study, children who were simultaneous bilinguals of English and Dutch outperformed Dutch children schooled in an English immersion programme on their perception of difficult English contrasts. These findings demonstrate that receiving L2 input solely during school hours, may not be enough to develop native-like phoneme perception abilities. From this data, it can be concluded that children who receive L2 input, both at home and during school hours, are more likely to reach native-like phonetic perception levels in an L2, and develop a better mastery of the phonetics of both their languages. In Chapter 2 we deliver our memory card phonetic trainer to a highly bilingual population who are exposed to Spanish and Catalan on a daily basis, in school. Some children are also exposed to both languages at home and all of them are learning English as an additional language. In our experiment, we propose that those children who speak both Spanish and Catalan at home have better knowledge of both of their languages. We use this distinction in degree of bilingualism to observe its effect on L2 phoneme learning.

1.4. Second-language Phonetic Learning in Adulthood

Second language phonetic learning in adulthood is more challenging than in infancy and childhood. Despite the passage of time, memory traces of an individual's first language, learned early in life, remain detectable in how sequential bilinguals process speech. This well-established phenomenon has been replicated by various research teams across multiple countries and several theories have explored the connections between the L1 and L2. These have mostly focused on the perceptual representations of phonological categories in both languages, to shed light on how initial language exposure shapes long-term phonological abilities. Although the current thesis does not directly test the predictions of any specific theoretical framework, the major theories of L2 speech learning will be explained here to contextualize the research and demonstrate how the findings of this thesis can contribute to and enhance existing knowledge in the field.

One of the theories that has received the most attention in L2 phonetic and phonological learning has been Best’s Perceptual Assimilation Model (PAM, 1995). The PAM explains L2 phoneme perception by means of a “phonological space” where sounds from the L1 are organized based on their similarities and differences in how they are produced. When naïve listeners hear an L2 phoneme, they will try to match the unfamiliar sound to the closest sounds in their native language, a process termed perceptual assimilation. If the L2 sounds resemble those of the L1, in terms of their gestural patterns (i.e., movements and positions of the tongue, lips and vocal folds) the non-native sounds will be easily categorised. However, if the L2 sounds are very different, they might be hard to categorize or to be recognized as speech sounds (Best, 1995).

Best and Tyler (2007), expanded the PAM to predict how L2 learners perceive speech, resulting in the PAM-L2 model. They identified various patterns of cross-language assimilation at both the phonetic level (including context-dependent and dialectal variations of the same phoneme) and the phonological level (such as lexical minimal pairs i.e., words differing in meaning due to a single phonetic contrast). According to the PAM-L2, when contrasting L2 phonetic segments can be categorized as examples of L1 phonological categories, three patterns of perceptual assimilation predict how difficult it will be for naïve listeners to discriminate between them. If the contrasting L2 sounds are perceived as belonging to separate L1 categories, discrimination is expected to be excellent (Two Category pattern). For example, German listeners are relatively good at discriminating the English /i/-/ɪ/ contrast (Bohn & Flege, 1990; Flege et al., 1997), most likely because German has a similar /i/-/ɪ/ distinction. However, Spanish listeners find this English contrast particularly difficult (Escudero & Boersma, 2004) because they only have one /i/ vowel in that acoustic region. This Spanish assimilation pattern is an example of the Single Category pattern whereby both segments are perceived as equally good examples of a single L1 category, making discrimination most challenging. If both segments are categorized under a single L1 category but differ in how well they fit that category, discrimination will be easier (Category Goodness pattern).

It could also be the case that at least one of the sounds in the L2 contrast cannot be categorised as an example of an L1 category. If one sound in the L2 contrast corresponds to a known L1 phoneme while the other does not (Uncategorized-Categorized assimilation) discrimination is expected to be good. For instance, in the Farsi voiceless uvular-velar contrast /q/-/k/, /k/ is assimilated into its English counterpart but /q/ does not exist in English and is not categorized, leading to adequate discrimination (e.g., Polka, 1992). If an L2 contrast does not align with any contrast in the L1, (Uncategorized-Uncategorized assimilation), the ability to distinguish between the sounds can differ depending on the specific contrast in question. For example, in Werker and Tees' (1983) hindi dental-retroflex stop contrast / t̪a/-/ ʈa/, discrimination accuracy was quite low. Lastly, when neither sound in the L2 pair matches is perceived as a speech sound, (Non-Assimilable) discrimination is typically very strong. This is because listeners treat both sounds as entirely unfamiliar, often perceiving them as non-speech. For instance, English speakers hearing isiZulu click consonants (Best et al., 1988). Overall, the PAM proposes that L2 segments are integrated into L1 categories with varying accuracy, depending on their articulatory-phonetic features or phonological roles, such as the range of pronunciations they can take and the contexts in which they occur (e.g., French [r] has more variation in pronunciation and depends more on context than English [r]). The patterns of phonetic and phonological assimilation, along with the functional load (i.e., how many word pairs a phoneme distinguishes) in the L2, collectively influence the likelihood that an L2 contrast will become perceptually distinct with increased L2 exposure.

The PAM, like the other theories, was developed using evidence from behavioural studies. Although they are very informative, a behavioural studies do not allow us to understand where the differences in phoneme perception lie between speakers of different languages. That is, whether they occur at early stages of processing or whether they involve higher-order processing of the speech stimuli. Chapter 4 of the current thesis provides an

EEG tool to explore cross-language vowel perception in English native speakers and Spanish learners of English. We use EEG as it is better suited for exploring early processes of acoustic processing than behavioural tasks, providing a very precise reading of the brain's response to speech sounds. The main aim of the study was to further refine an EEG tool that would allow us to test phonetic perception of the whole vowel space and would demonstrate effects of participant language background. In our study, some of the vowels presented to the listeners sounded like native vowels whilst some did not. We also collected perceptual judgements of these target vowels to observe their degree of assimilation into native vowel categories in English and Spanish. Therefore, Chapter 4 compares EEG responses with behavioural data and evaluates how well this method captures cross-language perceptual differences.

Other theories of L2-speech learning have explained this process of perceptual assimilation of phonemes in different ways. Kuhl's Native Language Magnet (NLM, 1993) states that the perceptual space has an inherent structure, where certain phonetic differences, defined by their acoustic properties, are easily distinguishable. For instance, /æ/ and /ɑ/ have greater spectral differences compared to /ʌ/ and /ɑ/ and hence, /æ-ɑ/ is easier to discriminate (Shafer et al., 2021). Moreover, the point vowels /i/, /a/, and /u/ are highly distinct in their acoustic and articulatory properties, making them easier to distinguish for both native and non-native listeners (Friedrichs et al., 2017). Under the NLM framework, as infants are exposed to their native language, these easily discriminated phonetic differences, known as "hotspots," act as magnets, attracting other phonemes and forming phonetic categories. This process leads to a reorganization of the infant's brain, focusing on the specific sounds of their L1. As a result, their perceptual systems become *warped* and frequently heard sounds emerge as prototypes of the phonetic categories, making native sounds more distinct and non-native sounds, harder to distinguish (Iverson & Kuhl, 1995; Kuhl et al., 2008). While the phenomenon described in the NLM model closely resembles the concept of perceptual assimilation in the PAM, the two frameworks differ in emphasis. The NLM does not focus explicitly on cross-linguistic perception; instead, it proposes that once L1 phonological categories are established, they shape how listeners perceive new speech sounds, making it harder to process them based on their raw acoustic properties.

Similarly to the NLM, the Automatic Selective Perception (ASP) model of L2 speech perception (Strange, 2011) describes the effects of the L1 on L2 perception as the result of perceptual specialization for the native language. According to the ASP, native language acquisition involves the development of selective perception routines (SPRs) that allow listeners to process speech automatically and enhance speech perception in adverse conditions. These routines guide listeners to focus on a limited set of acoustic features within the speech stream, filtering out others that could also be noticed. Unlike models such as PAM-L2 and NLM, which emphasize either cross-language category mapping or prototype formation, ASP centers on how learners assign importance to different auditory cues. The ASP suggests that language experience fine-tunes listeners' attention to linguistically relevant acoustic cues, making it harder to perceive distinctions not used in the native language. This attentional bias, according to ASP, plays a key role in the difficulty adults face when learning non-native phonemic contrasts.

High Variability Phonetic Training (HVPT), explained in Chapter 2 of this thesis, was developed with these principles in mind. By exposing learners to a wide range of talkers and tokens, HVPT aims to redirect learners' attention to the acoustic dimensions that are crucial for distinguishing non-native contrasts, effectively countering the perceptual narrowing predicted by the ASP model. In HVPT paradigms, learners are trained to recognize specific sounds, which can include both speech segments,like vowels and consonants, and suprasegmentals, such as lexical tone, produced by various speakers in different phonetic contexts. In a foundational study by (Logan et al., 1991), Japanese speakers learning English as an L2 listened to minimal pairs of words that contrasted /l/ and /r/ sounds (e.g., lead vs. read). They identified the sounds they heard using a two-choice identification task and received immediate corrective feedback. This training significantly improved their perception accuracy from 78.1% before training to 85.9% after training. Subsequent research has shown that similar training improves the perception accuracy of many other L2 sounds, including vowels (Lambacher et al., 2005; Thomson, 2012), stops (Flege, 1995), fricatives (Lengeris & Nicolaidis, 2015), tones (Wang et al., 2003), and syllable structures (Huensch & Tremblay, 2015).

In particular, Chapter 2 explores the effects of age and bilingualism on HVPT performance. To test the influence of language learning experience on HVPT performance, Tremblay and Sabourin (2012) trained mono-, bi- and multilingual speakers on the voiceless aspirated dental/retroflex stop contrast /t^h - $ʈ^h$/. All their subjects were native speakers of English; the bilinguals' additional language was French, which they learned between the ages of 3 and 10 years. Besides English and French, the multilinguals had learned an additional language from birth or between the ages of 2 to 28 years. Results on an AX discrimination task showed no group differences on the pre-training test but an increased accuracy in the multilinguals over the monolinguals, at post-test. Moreover, both the multilingual and bilingual groups were more accurate than a control group that had received no training. The authors took these results as an indication of increased perceptual sensitivities and greater phonetic learning abilities in multilingual and bilingual speakers. Moreover, Spinu et al., (2018) reported better production accuracy of glottal stops in Canadian bilingual speakers, compared to monolingual speakers, after being exposed to the Sussex English accent. Nevertheless, training studies on bi- and multilingual speakers remain scarce and it is difficult to gauge the replicability of this finding. Nevertheless, training studies involving bi- and multilingual speakers remain limited, making it difficult to assess the replicability of this finding. The training method presented in Chapter 2 collects perceptual data in an engaging, online format, offering a practical tool for future research to explore these questions more systematically.

In contrast to the L2-speech learning theories described above, Flege's Speech Learning Model (SLM) (1995; Flege et al., 2003) examines not only, L2 perception but also L2 production. Regarding L2 perception, the SLM's main focus is on the development of L2 phonological categories that evolve with L2 exposure. The SLM, similar to the PAM, predicts that if contrasting L2 sounds are assimilated into the same L1 category, it will be challenging to discriminate between them, as well as to distinguish the L1 sounds from the L2 sound. For instance, the southern British English /i- ɪ/ contrast being difficult for Spanish speakers, as there is only one i vowel in that acoustic region in Spanish. The main difference, regarding phonetic perception, between the SLM and the other three theories is the SLM's prediction of bidirectional cross-linguistic influence. Meaning that, as much as the L1 influences L2 perception, the L2 can also impact L1 phonetic perception.

The revised SLM-r (Flege & Bohn, 2021) further elaborates on these principles by emphasizing the dynamic nature of phonetic category formation and adjustment. It suggests that L2 phonetic categories are not static but continuously evolve with ongoing L2 exposure and use. This evolution is influenced by factors such as age of acquisition, amount of L2 use, and the quality of L2 input. The revised model also highlights the role of individual differences in cognitive and perceptual abilities in shaping the interaction between L1 and L2 phonetic systems. Adding to this, in the Auditory Precision Hypothesis-L2, Saito et al., (2022) suggest that domain-general auditory processing abilities contribute towards near-native L2 speech perception. Meaning that learners who are better able to tune into phoneme acoustics should acquire non-native speech categories with greater ease. In their study Saito et al., (2022) explored the relationship between sensitivity to different auditory features in non-verbal sounds (i.e., F3 frequency, F2 frequency, and duration), and proficiency in both perceiving and producing the English /r-l/ contrast. Their results showed that although the presence or absence of L2 immersion had the biggest influence on /r-l/ proficiency, performance was also linked to sensitivity to key auditory cues. In particular, greater awareness of F3 changes, which serve as the most reliable L2 cue, and sensitivity to F2 shifts, which are less reliable but already familiar, played a role. These findings suggest that, those learners who are better able to combine native-like and non-native like strategies for phoneme discrimination in their learning process, will be the most successful.

Other factors, such as the age of acquisition of the L2, the L1:L2 usage ratio, and language dominance have been shown to impact the evolution of L2 speech perception. The Age of acquisition (AoA) represents the age at which a person first encounters the language that will become their L2. This initial exposure is considered the start of L2 learning and it significantly influences the development of phonetic categories, with the resulting interaction between the L1 and L2 having a long-term impact on phonetic perception. Flege (1993) conducted a study with Chinese participants to examine their perception of the English stop consonants /t/ and /d/ in word-final positions. This contrast is challenging for Chinese listeners because their native language lacks word-final stop voicing distinctions. In English, native speakers tend to lengthen the vowels before /d/ but not /t/, which influences their perception of ambiguous word-final stops as voiced (i.e., /d/) when preceded by longer vowels. Flege compared two groups of Chinese-English bilinguals: early bilinguals who arrived in the United States by age ten, and late bilinguals who were regularly exposed to English only after arriving in the United States. Participants were presented with a continuum of vowel durations designed to bias their perception of the final consonant as /d/ or /t/. Both groups showed sensitivity to vowel length in determining the voicing of the final stop. However, late bilinguals exhibited a weaker effect of vowel length compared to early bilinguals and native English speakers, indicating that their later age of acquisition constrained their ability to perceive the relationship between vowel length and word-final stop voicing in English.

Flege et al., (1999) conducted a study with Italian-English bilinguals to assess their perception of various English vowel contrasts. The participants were divided into three groups based on their age of acquisition: early (around age seven), mid (around age fourteen), and late (around age nineteen). They were presented with vowel contrasts from their L1, Italian, their L2, English, and cross-language contrasts between Italian and English. The results indicated that as the age of acquisition increased, the accuracy of perceiving L2 vowels decreased, suggesting that an older age of acquisition negatively impacts the ability to accurately perceive L2 vowel sounds. Recent research has consistently shown that AoA significantly impacts learners' speech perception abilities (Birdsong, 2018). Moreover, it has been suggested that AoA is the main factor influencing non-native level proficiency in a second language (Bylund et al., 2021).

Besides AoA, the proportion of L1 and L2 use will also influence bilinguals' speech perception. Bilinguals utilize their languages in various social settings, with different conversation partners, and for multiple communicative purposes. For example, interacting with monolingual family members at home will differ from participating in a bilingual classroom, which will also differ from engaging in casual conversations with friends. The ratio of L1 to L2 usage provides an overall estimate of the time a person spends communicating in their first and second languages. Generally, more frequent use of a language is associated with higher proficiency in that language, which affects speech perception. Over the medium to long term, shifts in language use patterns are known to impact how bilingual listeners perceive speech. To systematically investigate this, studies often control for age of acquisition by comparing bilinguals who learned their languages at similar ages but have different L1:L2 usage ratios.

In a large study of L1:L2 usage ratio, Flege and MacKay, (2004) conducted four experiments to present English vowel contrasts to Italian-English bilinguals. In the first two experiments, Italian native speakers who had recently arrived in Canada (less than three months) struggled to distinguish the English vowel contrasts /ɒ/–/ʌ/, /ɛ/–/æ/, and /i/–/ɪ/, often categorizing both English vowels as the same Italian vowel. The researchers then explored the effects of age of acquisition and the amount of L1 Italian use on the discrimination of English vowels in two additional experiments. Experiments 3 and 4 compared Italian-English bilinguals who had lived in Canada for several years to Canadian English monolinguals. The bilinguals were divided based on their age of arrival in Canada and their daily use of L1 Italian, ranging from low (less than 15% of the time) to high (more than 25% of the time). This resulted in four groups: early low, early high, late low, and late high. These groups were tested on their ability to discriminate the English vowel contrasts /ɛ/–/æ/, /ɐ/–/ʌ/, and /i/–/ɪ/. The results from Experiments 3 and 4 collectively showed that early bilinguals were more proficient at distinguishing English vowel contrasts compared to late bilinguals. Additionally, bilinguals who used their native Italian less often were better at discriminating English vowels than those who used Italian more frequently. The study also found an interaction between age of acquisition and the frequency of L1 use. Early bilinguals who rarely used Italian performed similarly to native English speakers in perceiving English vowels, while those who used Italian more frequently differed from English monolinguals. These findings indicate that although age of acquisition influences speech perception, its effects are moderated by the ratio of L1 to L2 usage. Early bilinguals who frequently use their L1 are more likely to experience interference from their native language when perceiving speech in their second language.

The ratio of L1 to L2 use significantly impacts the relative strength of a bilingual's L1 and L2. Over time, this usage ratio influences language dominance. It was once believed that true bilinguals must master both languages equally, but research suggests that few bilinguals achieve equal proficiency in both languages (Grosjean, 2015). Typically, bilinguals have a dominant language, which is the one they use most frequently and feel most comfortable using. However, measuring language dominance is challenging due to the lack of consensus on its definition (Flege et al., 2002; Grosjean, 1998). Most studies show a processing advantage for the L1, even among fluent bilinguals who acquired their L2 early and use both languages daily. However, many studies have primarily involved L1-dominant bilinguals, which can skew results. Some researchers propose that bilinguals who extensively use their L2 may become L2-dominant and better suppress L1 interference when perceiving L2 speech (Flege et al., 2002). Although this hypothesis has been tested, the evidence remains mixed, indicating a need for further research.

Amengual and Chamorro (2015) investigated how language dominance affects speech perception by presenting Galician mid vowel contrasts (/e/–/ɛ/ and /o/–/ɔ/) to fifty-four proficient Spanish-Galician early bilinguals. Language dominance was measured using the (Birdsong et al., 2012), which considers factors like age of acquisition, L1:L2 usage ratio, competence, and language attitudes. The study found that Galician-dominant bilinguals, who were exposed to Galician earlier, used it more frequently, and spoke it more natively, showed strong categorical perception of Galician mid-vowel contrasts. In contrast, Spanish-dominant bilinguals struggled to differentiate these vowel sounds. This suggests that language dominance influences phonological organization, with Galician-dominant bilinguals forming distinct vowel categories, while Spanish-dominant bilinguals used merged categories. However, the study also noted that there are conflicting findings that limit the generalizability of the conclusion that language dominance is a key factor in shaping bilingual speech perception.

Additionally, Casillas, (2015) examined how different groups of bilinguals and monolinguals categorize the Southwestern American English tense-lax high front vowel distinction (/i/–/ɪ/). The study involved ten English monolinguals, ten Spanish-English late bilinguals (who learned English as adults and were Spanish-dominant), and ten Spanish-

English early bilinguals (who learned Spanish in early childhood but were now English-dominant). The results showed that English monolinguals had the most distinct category boundary for the vowel contrast, followed by English-dominant bilinguals, and then Spanish-dominant bilinguals. The groups also differed in the acoustic cues they used for categorization. English monolinguals and English-dominant bilinguals relied more on vowel spectrum properties, while Spanish-dominant bilinguals depended more on vowel duration. Interestingly, English-dominant bilinguals still showed some sensitivity to vowel duration, a trait typical of native Spanish speakers, despite their early acquisition of English and lack of current use of Spanish. These findings suggest that the influence of the L1 can persist in shaping the perception of L2 speech sounds, even after many years of exclusive use of the L2.

Taken together, these findings suggest that L2 perceptual abilities are shaped not only by the age at which the language is acquired, but also by the quantity and quality of L2 exposure, and the relative dominance of each language. However, once again, most of this evidence relies on behavioural measures, which do not reveal the underlying neural mechanisms driving perceptual differences across individuals. This is where the EEG study in Chapter 4 offers a valuable contribution. By using the EEG tool to compare early neural responses to vowel sounds in language learners, we can provide insights into how L1 and L2 experience shape low-level auditory processing. Because EEG allows us to isolate early, automatic stages of speech processing, it can help disentangle the effects of factors like AoA and language use from later, more strategic or task-driven processes. As such, the paradigm introduced in Chapter 4 represents a promising method for deepening our understanding of how L1 and L2 experience interact to influence phonetic perception at both neural and behavioural levels.

1.5. The current thesis

As mentioned above, the current thesis is divided into three experimental chapters. Chapter 2 reports a phonetic training study which used a computerised memory-card game (similar to the game of Concentration or Memory) to observe effects of age and bilingualism in L2 phoneme discrimination. In their English language classes, Spanish-Catalan children (6-12 years) were trained on the Southern British English vowel set with an audio version of the Memory game, in which they saw a grid of face-down cards, and clicked on card pairs to find the matching pairs. A match consists of the same word spoken by two different speakers, chosen from cards representing words within the same cluster of minimal pairs. A three-alternative forced-choice oddity discrimination task was used to track phonetic learning in a pre- and a post-training test session. Results showed that the game enhanced children's perception of the Southern British English vowel set and hence, the memory-card game was a suitable platform for training children. Although this task was designed to be appropriate for young children, older children learned more, especially those exposed to multiple languages at home. These results are discussed in Chapter 2, with regards to children's perceptual abilities and methods to improve child L2 phonetic learning in diverse linguistic contexts.

Chapter 3 tests a novel tool for examining perception of multiple phonetic contrasts in a time-effective way. For this purpose, a vowel-pair oddball paradigm with 19 IPA vowels is presented to British English adults to elicit Mismatch Negativity (MMN) responses in EEG recordings. Neural responses are analysed using a traditional event-related potential (ERP) approach and by fitting multiple Temporal Response Functions (mTRFs; Crosse et al., 2016) in a backwards model. Results revealed that, the oddball paradigm elicited reliable MMNs with peaks between 100-250ms after onset of the deviant stimulus. Moreover, the mTRF analysis method successfully extracted reliable MMN responses from the EEG data after a single deviant instance per deviant type tested (i.e., 19 different IPA vowels). This is a remarkable finding since the mTRF methodological framework for MMN allows for inclusion of more deviant types in the oddball paradigm, allowing for the examination of perception of multiple contrasts. Nevertheless, the recorded MMNs did not show reliable components of phonological processing. Therefore, these results are discussed regarding the literature on MMNs as neural indices of phonetic and phonological perception as well as describing how the MMN tool could provide a suitable platform for exploring the development of phonetic learning throughout the lifespan.

Chapter 4 explores cross-language vowel perception using a modified version of the oddball paradigm used in Chapter 3. The main aim of this last experiment is to obtain MMNs that can be used as indices of neural vowel perception to map out participants' perceptual vowel spaces. Recordings of EEG data were obtained from British English (N=18) and Spanish (N=17) subjects and given the success of the analytical methodology in Chapter 3, the MMNs were, once again, analysed with backwards multivariate Temporal Response Functions (mTRFs; Crosse et al., 2016). The back-projected data revealed reliable MMNs for the 19 IPA vowels, each showing different magnitudes. Using these MMN magnitudes, the pre-attentive vowel spaces were mapped for the English and Spanish participants and reliable cross- language effects were found. To explore the relationship between MMNs and behavioural identification responses, a vowel identification and category goodness task was delivered to participants, in their native languages. Results showed that, Spanish speakers assimilated the IPA vowels more strongly into native categories than English participants and increased assimilation scores led to larger-amplitude MMNs. These results demonstrate that MMNs can be used to create perceptual maps for IPA vowels, in a way that reflects perceptual differences across English and Spanish speakers. Therefore, this MMN framework could be applied to more bilingual populations across different age groups to observe how vowel categorization evolves over the lifespan, enhancing our understanding of phonetic/phonological development. These three investigations will be discussed in Chapter 5, which will provide a general discussion of the main findings of the current thesis.

2. Memory-Card Phonetic Training of English Vowels for Spanish-Catalan Children.

2.1. Introduction

Learning to speak a second language (L2), with a native-like accent, becomes increasingly harder with age. For instance, research by Flege et al., (1999) among Italian expats in Canada revealed that, with time, children eventually developed a native-like English accent, in contrast to adults who migrated later in life. Two main factors play into this: a decline in neuroplasticity and differences in the language input that children and adults receive (Flege, 2008; Herschensohn, 2007). It is well known that the neuroplasticity of the child brain allows them to acquire new skills with ease, such as language. On top of this, children are more likely to assimilate into their new culture and make connections with native speakers of their L2, whereas adults tend to maintain ties to their home-country communities (Zhang et al., 2018). To explore the influence of neuroplasticity in L2 speech learning whilst controlling for speech input, researchers can use High Variability Phonetic Training (HVPT) programs, that deliver the same intensive L2 phoneme exposure to all participants (e.g., Logan et al., 1991). The aim of these programs is to familiarise listeners with the acoustic properties in the non-native phonemes, so they learn to extract cues for L2 phoneme identification and discrimination. Surprisingly findings from HVPT programs, do not show consistent advantages in phonetic learning in children over adults (Heeren & Schouten, 2008, 2010; Wang & Kuhl, 2003).

The present investigation uses a novel L2 phoneme training technique to explore phonetic learning amongst primary-school children as a function of age. HVPT programs were initially designed for adult learners, but children learn from them too. For instance, Wang and Kuhl (2003) trained English-speaking children (ages 6, 10, and 14) and young adults in Mandarin tones. Following six training sessions, all groups demonstrated improved tone identification, with no discernible differences in the levels of improvement between age groups. Some studies have reported better performance in adults, like Heeren and Schouten (2008, 2010), who trained both Dutch 12-year-olds and adults in the Finnish /t/-/tː/ contrast. They reported that after five training sessions, both groups showed learning, but adults exhibited greater enhancement in discrimination abilities compared to children. Additionally, the effects of plasticity in HVPT have been shown to interact with the length of training, as studies with more training sessions, have reported better learning in younger groups. Shinohara and Iverson (2021) trained Japanese-speaking children, adolescents, and adults on the /r/-/l/ contrast across 10 training sessions and found that the younger groups improved their /r/-/l/ perception more than adults, with adolescents outperforming children. Giannakopoulou et al., (2013) employed an audio-visual approach to train 8-year-old Greek children and adults on the tense-lax /ɪ/-/i/ contrast over 10 sessions. Initially, children demonstrated greater improvement than adults in the /ɪ/-/i/ contrast. However, in a subsequent study where /ɪ/-/i/ were linked to images instead of graphemes, Giannakopoulou et al. (2017) found no evidence of Greek children outperforming adults demonstrating that, task complexity also impacts the degree of learning in younger listeners.

Moreover, performance on HVPT is seen to improve throughout childhood, as older children tend to outperform their younger peers. Kasisopa et al. (2018) trained Thai-, English- and Arabic-speaking 6- and 8-year-olds on the mandarin tones. They found a minimal improvement in tone perception in the 6-year olds and the 8-year olds, showed significantly greater improvement in tone identification. Shinohara and Iverson (2013) trained 6- to 8-year old Japanese-speaking children on the English /r/-/l/ contrast and found that the older participants improved their /r/-/l/ perception more than the younger ones. Moreover, Brekelmans (2020) found that although Dutch 7- to 8-year-olds learned English

L2 vowels during training, they did not show improvement in a post-training test of L2 vowel perception. However, in her study, a significant improvement was seen in some of the post-training tests in her 11-12-year-old participant group. Overall, these findings point towards a potential disadvantage in 6– to 7-year-olds in HVPT as compared to children at the age of 8 years and above. Nevertheless, phoneme learning in younger children may have been influenced by task characteristics, as high cognitive demands in phoneme identification tasks may disproportionately affect younger learners, who typically have more limited working memory resources (Gathercole et al., 2004).

We created a new training method tailored for children as young as six, using a familiar card-matching game format, commonly known as *Memory* or *Concentration* (Iverson et al., 2023). The classic card game involves matching pairs of images by turning over two cards at a time from a face-down grid. If the cards do not match, they are turned back over, and players rely on their memory of previously revealed cards in subsequent turns. While both adults and children aged 6 to 8 can perform this task similarly, six-year-olds tend to be somewhat slower (Krøjgaard et al., 2019). In our version, the cards were linked to minimal pair words spoken by multiple talkers. Each card in the game is associated with a recorded word. When a player flips a card, the corresponding word is played aloud. If the player flips two cards that match (i.e., they are linked to the same word), the pair is removed from the grid, otherwise, they are turned face-down again. We chose to adapt this game because it is familiar and enjoyable for many children, and it allows them to play based on sound alone, without requiring reading skills or knowledge of word meanings. Functionally, it serves as a gamified version of a category discrimination task, where listeners focus on phonetic category distinctions across different speakers (Gottfried, 1984; Shinohara & Iverson, 2018), although the game imposes a greater memory demand.

Although the task demands are likely the most important reason why younger children perform less well in HVPT, being at an earlier stage in their linguistic and cognitive development may make 6-year olds less able to learn from these training protocols. Burnham (2003) found that perception of native contrasts in 4-, 6- and 8-year olds improved gradually with age whereas perception of non-native contrasts showed a U-shaped developmental curve with lowest thresholds at the age of 6 years. Moreover, the degree of speech-specific perception of the participants was linked to their early reading abilities suggesting that, as children begin learning to read, the process of mapping phonemes to graphemes may reinforce native language phoneme categories and reduce sensitivity to non-native contrasts. This would make L2 phonetic learning much harder at 6 years of age. On top of this, longer experience with language could help older children benefit more from the variability in the training input. Children's categorisation strategies are not adult-like until after 12 years of age and they are less consistent in their categorization, than adults, when fewer cues are available (Hazan & Barrett, 2000). The variability in the HVPT input may make it harder for children to detect persistent cues held in L2 phonemes, hindering their learning. Moreover, previous studies have demonstrated that performance on HVPT is impacted by category knowledge in the learner's first and second-languages (Iverson et al., 2005; Iverson & Evans, 2009a), which in turn, develops with age. Altogether, this could explain why older children are consistently better in HVPT than their younger peers.

It is also possible that HVPT frameworks do not accurately reflect how phonetic learning occurs in real-world contexts. For instance, in these paradigms, participants get explicit trial-by-trial feedback on their responses, which is not as available in everyday communication. Recent studies have incorporated 'implicit learning' paradigms to study auditory category learning (Wade & Holt, 2005) and L2 speech learning (Lim & Holt, 2011; Saito et al., 2022). These multimodal tasks, such as shooting aliens or clay targets in response to specific sounds, engage working memory and visual processing alongside auditory perception. Importantly in these paradigms, learning takes place incidentally, without conscious awareness, and as a by-product of playing the game (see Saito et al., 2022). Although these studies show that adults can improve their perception of the trained stimuli (e.g., Japanese learners of English /r/ and /l/), no robust effects of generalisation to untrained stimuli have been found. Nevertheless, these paradigms resemble naturalistic L2 learning environments, where learners process auditory information while they focus on other goals.

In our memory card game, children are exposed to three different learning environments, and the need to attend to the auditory information varies in each. The easy symbols condition creates the most obvious context for implicit learning, as children can match the pairs focusing solely on the icons on the flipside of the cards (i.e., emojis, food items, sports equipment). As a result, children may not be aware of the words presented to them. In the hard symbols condition, less distinct icons (i.e., arrows pointing in different directions and similar geometric shapes) encourage children to attend to both symbols and words to complete the game. Lastly, in the no symbols condition, cards are identical visually and must be matched by sound alone. In this two latter conditions, children are likely to process the words to a greater extent. Nevertheless, the primary goal is to complete the game, not to learn the words. Moreover, the memory card game engages children in multimodal processing, prompting them to integrate auditory and visual cues. As a result, it creates a learning context that resembles real-life speech learning scenarios. The paradigm was not designed to evaluate implicit learning as such, but to provide a more ecologically valid training environment than traditional HVPT tasks, drawing on principles from implicit learning research.

The main aim of the current study was to examine the effects of age and bilingualism on phonetic learning in the memory card game. The current study took place in the classroom on the island of Majorca, whereby Spanish-Catalan children (6-12 years) were trained on the Southern British English (SBE) vowel set. All participants in the study are exposed to Spanish and Catalan daily in school but not all speak both languages at home – some speak other languages too. Therefore, they have varying levels of bilingualism. It has been proposed that bilingualism provides an advantage when acquiring an additional language (Abu-Rabia & Sanitsky, 2010). However, it remains unclear whether there is a bilingual advantage in perceiving non-native contrasts: some studies have found a bilingual advantage in discrimination of L2 speech segments (Enomoto, 1994; Rabinovitch & Parver, 1966) and other studies have shown that bilinguals do not outperform monolinguals (Davine et al., 1971; Gonzalez-Ardeo, 2001; Werker, 1986). Multilingual listeners demonstrate a learning advantage when trained on unfamiliar non-native contrasts (Tremblay and Sabourin, 2012), and this effect is influenced by the phonological similarity between their native and non-native languages (Antoniou et al., 2015). In a series of adult studies, Antoniou et al., (2015) found that Korean-English bilinguals outperformed Mandarin-English bilinguals and English monolinguals in discriminating phonemes in a Korean-like artificial language. However, the performance of the Mandarin-English bilinguals and English monolinguals did not differ. Kasisopa et al., (2018) found that speaking a tonal language facilitates learning of a new tone. In their study, English-Thai 8-year olds showed the most learning in mandarin tones, compared to their English-Arabic bilingual and their monolingual peers. Overall, if a larger phonetic inventory supports L2 learning (Iverson et al., 2005; Iverson and Evans, 2009), children with greater bilingual exposure might have an advantage due to their combined phonetic knowledge across both languages.

Spanish features a five-vowel system (/a, i, u, e, o/), while Majorcan Catalan has eight vowels (/a, ə, ɛ, e, i, ɔ, o, u/). Figure 2.1. below displays the average formant plots of three Spanish-Majorcan Catalan female speakers, illustrating key patterns in their vowel production, along with normalized average English vowels from our training set. The vowel systems of Catalan-Spanish bilinguals differ based on language dominance (Amengual, 2016; Simonet & Amengual, 2020), and the two languages share some overlapping vowel sounds. In particular, Majorcan Catalan has a denser vowel distribution in the lower-central area of the vowel space compared to Spanish, while both languages are fairly similar when it comes to high vowels. Previous research has shown that having a greater number of vowel categories can actually support the learning and discrimination of new contrasts, even though one might expect that fewer categories would leave more perceptual space for forming new ones (Iverson & Evans, 2009). Nonetheless, Spanish- Catalan bilinguals still need to acquire new phonemic categories to accurately perceive the more complex set of English vowels.

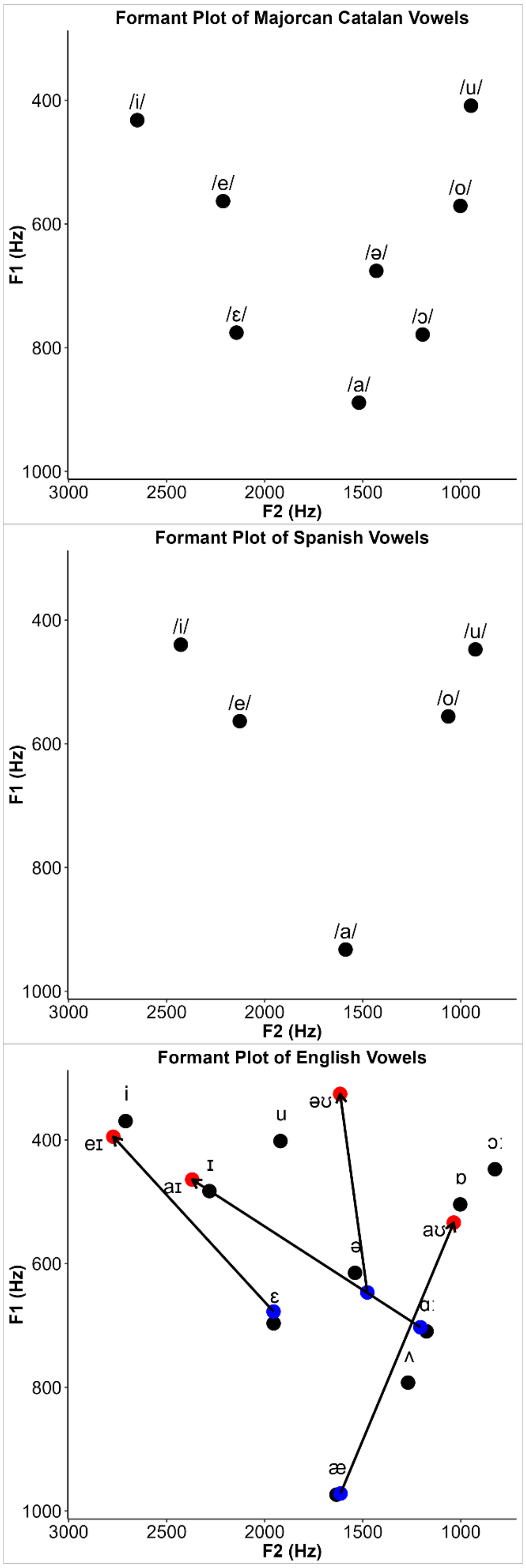

Figure 2.1. Formant plot illustrating the average vowel spaces of three Spanish-Catalan speakers. For reference, the average English vowels from the pre- and post-training experiment are also shown, with log-scaled average formants adjusted to align with the Spanish-Catalan vowel spaces. Arrows indicating the transition from the initial (blue) to the final (red) formant values represent diphthongs.

In this study, children aged 6 to 12 participated in memory card training sessions during their regular English lesson times. In a previous study we found that 6-year-olds lost interest in a three-alternative forced-choice oddity generalization task, as their performance declined over time (Iverson et al., 2023). To address this, the current study reduced the frequency of the oddity task and used it only as a pre- and post-training test of phonetic perception. Moreover, we expanded the set of animal characters used (e.g., bears, pigs, ladybugs, and frogs). In the oddity task, children heard three words spoken by different talkers, each paired with an animated animal, and had to identify which word was different (two were identical words from different speakers; the third was a different word). The cognitive requirements of the oddity task favour older children, with more mature working memory and phonological abilities. Nevertheless, performance on this task is not as affected by visual-spatial memory skills as the memory card game, and it offers a straightforward way of testing phonetic perception in a classroom setting. To assess generalization, two controls were added. First, two age groups (8–10-year-olds) were excluded from training but completed the pre- and post-tests for comparison with trained peers. Second, the tests included both trained and untrained stimuli to examine whether learning transferred beyond the items used in the memory game. The intervention spanned five 45-minute English lessons over two weeks: pre- and post-tests on the first and last days, and memory card games on the three days in between. Half of the games required children to focus on sound only, with identical ear symbols on all cards. The other half included visual symbols to be matched along with the sound. Of these, half used engaging, emoji-style icons (e.g., foods, faces), while the rest used simple, abstract symbols (e.g., arrows). There was no semantic link between the spoken words and the visual symbols, for instance, two cards that matched on the word *field* might also match on an unrelated symbol, like a left-pointing arrow.

2.2. Methods

2.2.1. Subjects

One hundred and three children (Mean $_{age}$ = 9;35 years, SD $_{age}$= 1.84 years) were recruited from two schools in the towns of Inca and Campanet, on Majorca. They were tested in their class groups in which all students were born in the same year: one class each aged 6-7, and two classes each aged 8-9, 9-10 and 11-12 year old. Of these children, 24 were designated as control subjects (two classes of students for which the total age 165 range was 8-10 years) and 79 received training (five classes of students). This testing occurred after the initial lockdowns of the Covid-19 pandemic, when there were still frequent absences from schools due to positive tests. Of the trained children, 63 children attended all three sessions of memory card training and out of those, 56 also were present for both the pre- and post-test tasks; only complete datasets were used in our statistical analyses because the exact sessions of our partial participants varied.

These children all spoke and heard Spanish and the Majorcan dialect of Catalan daily at school. Parental questionnaires revealed that 28 of the full cohort spoke only Spanish at home, 40 spoke only Catalan, 19 spoke Spanish and Catalan, and 16 spoke Spanish and another language at home (Arabic, Polish, Romanian, and English, with the English speaker's data not being included). In the 6-7 year-old classes, 15 children spoke only one language at home and 5 spoke two languages at home. In the 8-9 year-old classes, 16 children spoke only one language at home and 9 spoke two 10 languages at home. In the 9-10 year-old classes, 18 children spoke only one language at home and 10 spoke two languages at home. In the 11-12 year-old classes, 19 children spoke only one language at home and 11 spoke two languages at home.

2.2.2. Stimuli and apparatus

The games were written in our lab using JAVASCRIPT, PHP, and MYSQL and were designed to run within a web browser on the participant's computer. Children were tested in their classroom, using individual laptops and earphones.

The stimuli were natural recordings of Received Pronounciation English vowels used in previous vowel training experiments (Iverson et al., 2012; Iverson & Evans, 2009). All recordings had 44 100 16-bit samples/s but were compressed to MP3 format for faster internet delivery. There were 14 vowels arranged into clusters of three or four based on a hierarchical clustering analysis of second-language confusion matrices previously collected in our lab: / i, ɪ, aɪ, eɪ /, / ɛ, ɑ, a, ʌ /, / ɒ, əʊ, ɔ / and / u, aʊ, ɜ /. Response options on each pre/post-trial and competing minimal pair words in the memory card game were selected within these clusters (e.g., *field, filled, failed, filed*).

The training stimuli were spoken by two female and two male speakers, who produced 9 minimal pairs for each cluster (e.g., Ben-barn-ban-bun, field-filled-filed-failed, cod-code-chord, shoot-shout-shirt). The pre- and post-tests included these trained stimuli along with an additional set of untrained words (bVt) spoken by five female and five male speakers. Figure 2.1. illustrates the average formant frequencies of the vowels in the bVt words used in both tests, calculated based on the average formants from recordings of three female speakers.

2.2.3. Procedure

The training program involved a pre-training session in which children's perception of the Southern British English vowels was assessed, followed by three sessions of phonetic training, and it concluded with a post-training session in which SBE vowel perception was measured again, to observe changes from pre to post test performance.

At pre and post test, the children played the oddity task and in the three training sessions, they played the memory card game. Importantly, only the children in the experimental group, and not those in the control group, took part in the three memory-card training sessions. All children played the games during their English lessons and each testing session lasted around 45 minutes, including time for set-up activities and breaks. Children started and finished each task simultaneously, rather than ensuring every child completed the same number of trials or games. This meant that faster children played more games, while slower classmates were not rushed to keep pace.

At the beginning of each session, children got their own laptop and earphones and were told what task they would complete (whether they would complete the oddity task or the memory card game). For the oddity task, children were told that they would see three animals on the screen and that they should listen carefully to the words that the animals spoke. They were instructed to select the animal that said the odd word, using the cursor on the screen and to complete as many trials as they could before they were instructed to take a break. For the memory card game, they were asked to find the matching pairs on the grid, based on the words that were linked to each card and to complete as many blocks of the game as possible before the break. On top of this, the memory card game displayed visual feedback at the end of each round, indicating how long it took the participant to complete a game. This encouraged children to increase their speed in the memory card game.

To access the games, the children opened a generic link from their browser which directed them to the experimental online platform. In each session, the experimenter used a touchscreen tablet to control which task was available to the children. The experimenter could remotely start and stop tasks and manage transitions between activities in real time. The duration of each task was also controlled by the experimenter. When it was time for a break, a message was sent from the tablet to the children's laptop screen, prompting them to complete the current trial and then pause. This setup ensured consistent timing and coordination across sessions.

Children completed the oddity task in three six-minute rounds, with short breaks between rounds. On average, they completed 149 trials in the pre-test and 141 trials in the post-test. Each round included both trained and untrained stimuli. To keep the activity engaging, four types of cartoon animals (frogs, pigs, bears, and ladybugs) were used. Testing started with the same animal for each trial and progressed to featuring multiple animals within a single trial. All within-cluster pairs were presented (e.g., beat-bit, beat-bite, beat-bait, bit-bite, bit-bait, bite-bait; bet-Bart, bet-bat, bet-but, Bart-bat, Bart-but, bat-but; bot-boat, bot-bought, boat-bought; boot-bout, boot-Bert, bout-Bert), along with some cross-cluster pairs designed to assist children struggling to perceive differences (e.g., Bart-beat, Bart-boot, beat-boot). Correct answers were rewarded with the message "¡Muy bien!" and an animation where the animal stretched vertically and opened its mouth as if jumping. Incorrect answers triggered the message "Oh no..." with the animal shown against a red background. Figure 2.2 shows screenshots of correct trials (on the left column) and incorrect trials (on the right column) in the oddity task.

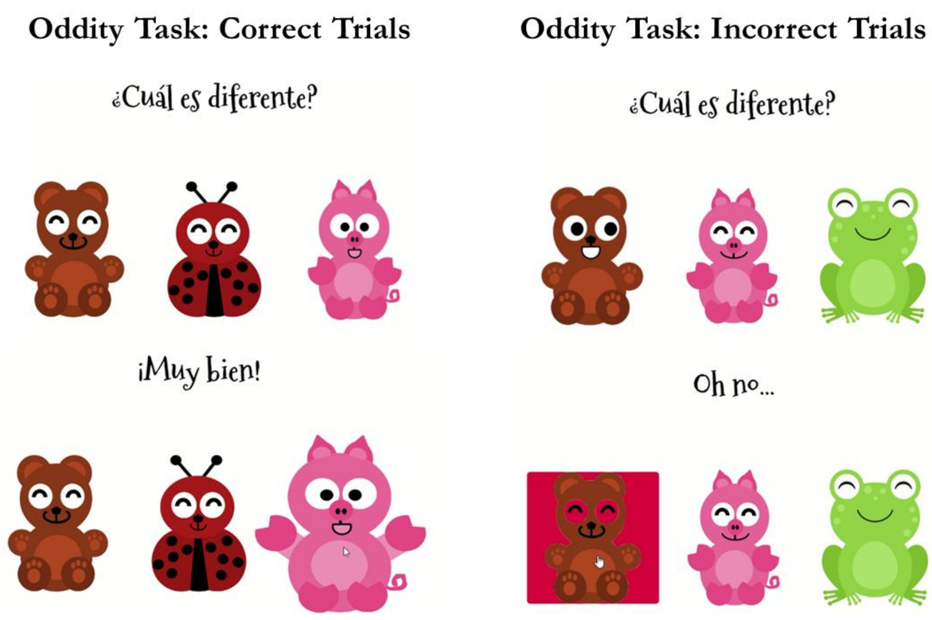

Figure 2.2 Screenshots of the oddity task. The left column shows, the three animated characters speaking, at the top. At the bottom, the pig is selected as the animal saying the odd word accompanied by positive feedback. The right column illustrates an incorrect trial of the same task.

In each MC game, 14 cards were displayed on the screen in a 3 x 5 grid with one missing position. Each card was linked to a different recording, comprising 2 talkers x 7 words (i.e., minimal pairs combined from three- and four-vowel clusters such as field, filled, filed, failed, cod, code, chord). The word clusters, speakers and position of the word in the grid were chosen randomly. Throughout the game, children chose pairs of cards and heard a word when each card was flipped over. Each game belonged to one of three conditions. In the no symbols condition there were identical markings on the underside of each card and matching was done purely by sound. In the easy symbols condition, the card pairs had matching symbols on their underside, which were semantically unrelated to the simultaneous word (i.e., icons representing sports equipment, food items or emojis). In the hard symbols condition, each pair was had symbols on their underside that were harder to tell apart from the competing pair symbols in the grid (i.e., arrows and geometric shapes pointing in different directions). The number of trials in each condition, was counterbalanced. In each session, children played the MC game for two rounds of approximately 10 minutes, with beaks in between rounds. Figure 2.3 displays two screenshots of the memory card game in the no-symbols condition: the left grid illustrates a correct attempt at matching the selected cards, while the grid on the right depicts an incorrect attempt.

Figure 2.3. Screenshots of the memory card game. The left grid demonstrates a correct match and the right grid, an incorrect one. In this block of the game, ear symbols were displayed on all cards so children had to focus on the audio cues to find matches.

2.3. Results

The statistical analyses were mixed models calculated using the lme4 package (Bates et al., 2015) in R. The CAR package (Fox & Weisberg, 2011) was used to obtain p-values within a type II analysis of deviance table. In the main analysis of responses in the oddity task, we also employed a stepwise procedure dropping non-significant factors (i.e., anova() comparisons, $p > 0.05$). All fixed effects are described below.

The first objective was to determine whether the training yielded significant training effects. Therefore, performance on the oddity task was compared between trained children and the group of age-matched, untrained, controls. We built a logistic mixed-model analysis with two fixed effects: pre vs. post and test group (experimental vs. control). By-subject, by-item and by-oddity-position (i.e., which of the three stimuli had a different word) were added as random intercepts. Figure 2.4. shows that proportion correct responses at post-test were higher in the experimental group than the control group and Table 1. shows the structure of the model and the values for the fixed effects. There was a significant interaction between pre/post-test and test group, $\chi^2(1) = 9.25$, $p = 0.002$, indicating greater performance improvement in the trained group. Additionally, there were main effects of pre/post-test, $\chi^2(1) = 7.75$, $p = 0.005$, with higher scores in the post-test overall, and of test group, $\chi^2(1) = 4.62$, $p = 0.032$, with the trained group achieving higher scores overall. These findings suggest a genuine improvement in phonetic category discrimination following training, despite children not being explicitly trained on this task. As a result, control subjects were excluded from further analyses of training effects.

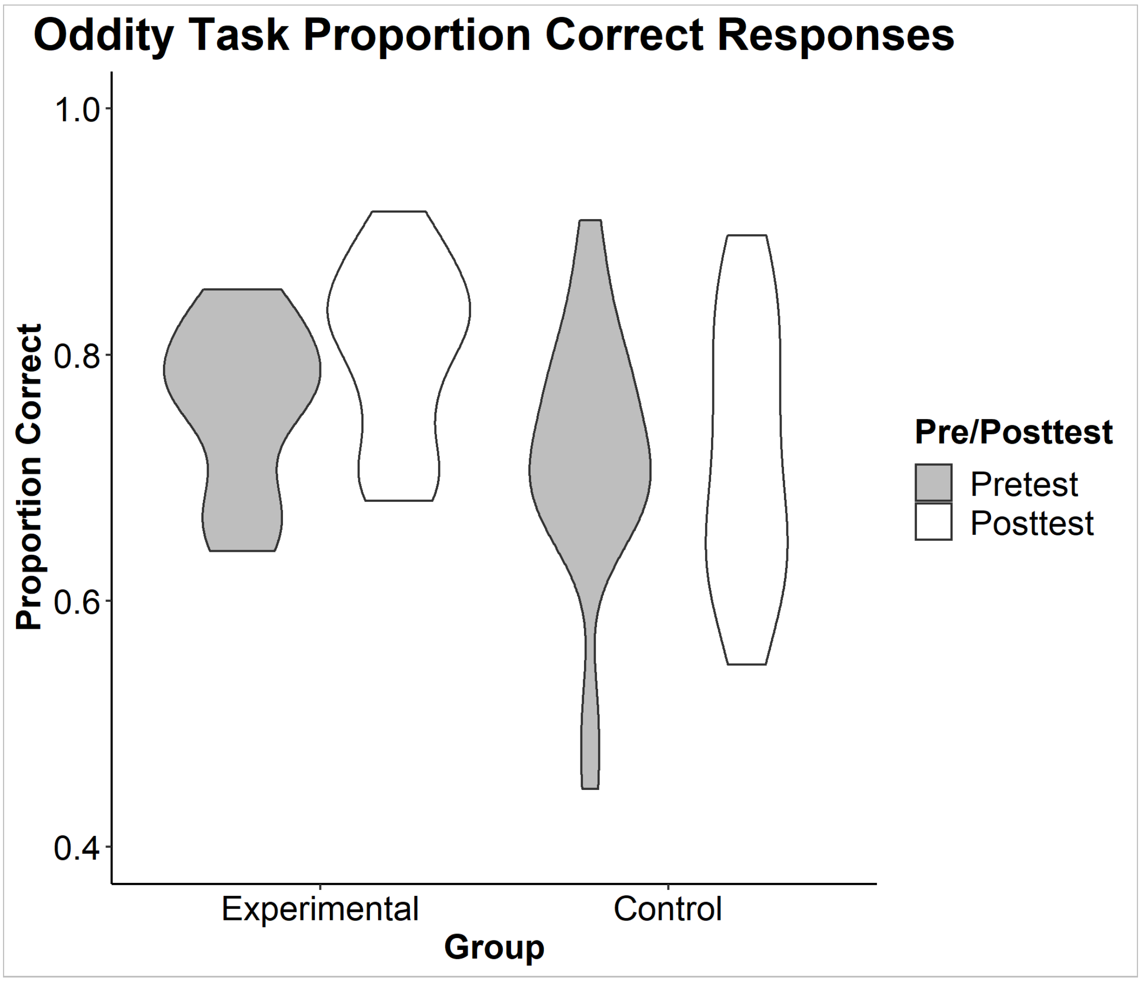

Figure 2.4. Violin plots illustrating the range of category discrimination accuracy for participants in the experimental and control groups, measured both before and after training. The experimental group demonstrated a significantly larger improvement in accuracy compared to the control group.

	Correct/Incorrect			
Fixed Effects	Estimate	SE	*z*	*p*
(Intercept)	**1.22**	**0.25**	**4.89**	**<0.001**
Pre/post	**0.27**	**0.06**	**4.12**	**<0.001**
Experimental/control	-0.21	0.16	-1.29	0.197
Pre/post: experimental/control	**-0.27**	**0.9**	**-3.04**	**<0.001**

Table 1. The GLMER model describes the influence of pre/post testing, between the experimental and control groups, on correct responses in the oddity task.

The oddity task was designed to assess generalization by including both, words that were trained in the memory card game and untrained words that participants had not encountered during training. To evaluate the extent of generalization, a second analysis compared improvements from pre- to post-test for trained versus untrained words, testing whether learning effects extended beyond the specific stimuli used in training. A logistic mixed-effects model was used, with fixed effects for trained vs. untrained words and pre vs. post-test fixed effects (see **Table 2.**). Random intercepts were included for subjects, items, and oddity positions. As shown in Figure 2.5., no significant interaction was found between pre/post-test and trained/untrained words, $\chi^2(1) = 1.42$, $p = 0.23$. Significant main effects were observed for pre/post-test, $\chi^2(1) = 47.88$, $p < 0.001$, indicating higher accuracy after training, and for trained/untrained words, $\chi^2(1) = 120.65$, $p < 0.001$, showing overall higher accuracy for trained words. Due to the absence of a significant interaction between trained/untrained words and training, this factor was excluded from subsequent analyses.

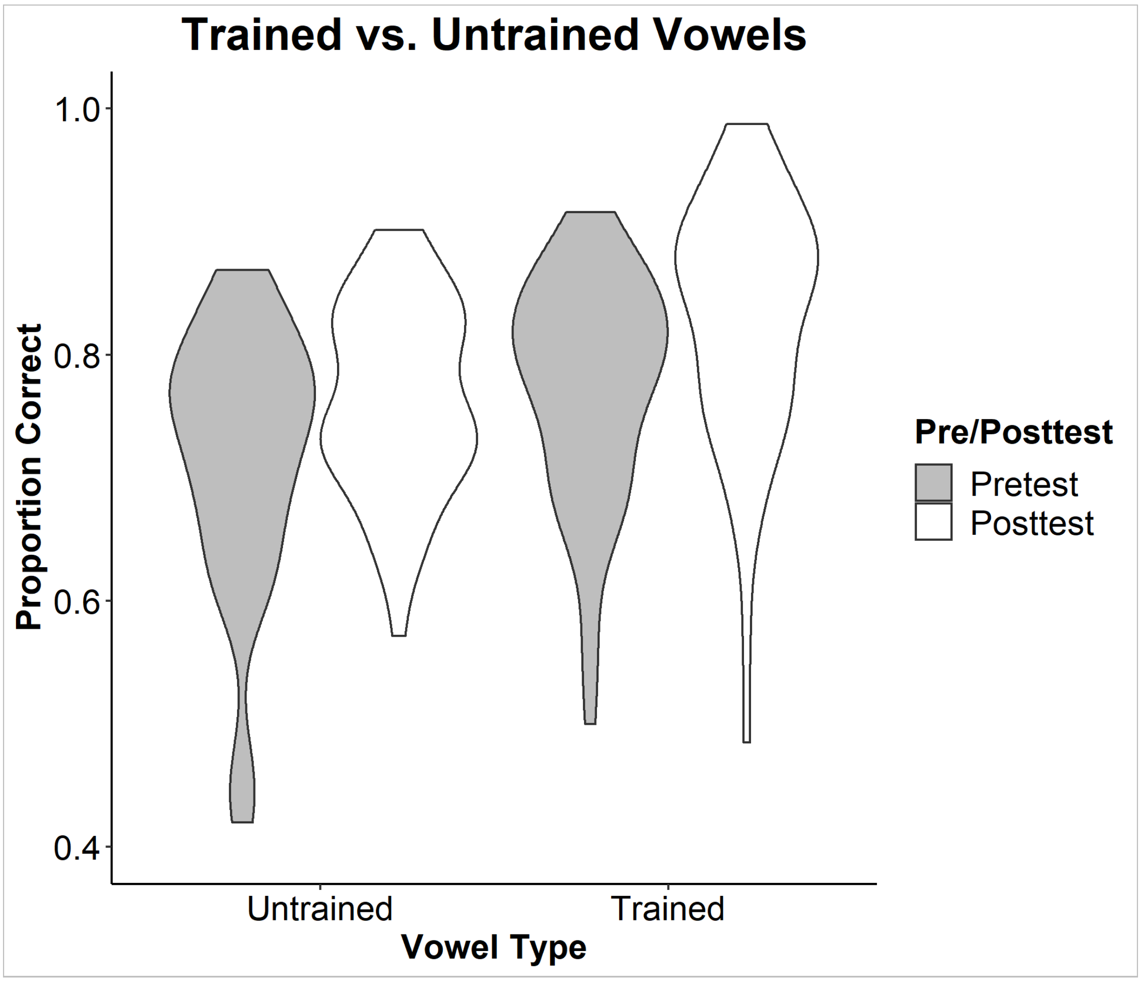

Figure 2.5. Violin plots showing the range of category discrimination accuracy for trained and untrained words, both before and after training. Post-test scores were significantly higher than pre-test scores, reflecting an improvement in performance following training. However, no significant interaction was observed between training and word type (trained/untrained).

	Correct/Incorrect			
Fixed Effects	Estimate	SE	*z*	*p*
(Intercept)	**0.93**	**0.24**	**3.94**	**<0.001**
Pre/post	**0.24**	**0.06**	**4.30**	**<0.001**
Trained/untrained	**0.40**	**0.06**	**7.32**	**<0.001**
Pre/post:trained/untrained	0.10	0.08	1.21	0.23

Table 2. The GLMER model describes the influence of pre/post testing on correct responses in the oddity task, within trained and untrained words.

The primary goal was to investigate the effects of age and bilingualism on learning. Participants spoke either one or two languages at home. Single-language homes used Spanish or Catalan, while two-language homes included Spanish and Catalan or Spanish paired with another language. To analyse these distinctions, three orthogonal dummy codes were created: number of home languages, Spanish versus Catalan for single-language homes, and Spanish-Catalan versus Spanish-other for two-language homes. Model comparison using the AIC criterion was employed to drop factors that did not contribute to the model.

For children from single-language homes, whether the language was Spanish or Catalan did not influence the model. Among children from two-language homes, the distinction between Spanish-Catalan and Spanish-other contributed to the model but did not interact with training, leading to its exclusion. The final model retained a language background categorical fixed effect with a one versus two home languages distinction. Age at testing was included as a linear fixed effect, and pre/post-test as a categorical fixed effect. The logistic analysis included random intercepts for items and subjects; random intercepts for oddity positions were excluded due to convergence issues. The model is summarised in **Table 3**.

	Correct/Incorrect			
Fixed Effects	Estimate	SE	*z*	*p*
(Intercept)	-0.05	0.42	-0.125	0.901
Pre/post	-0.20	0.22	-0.894	0.371
Home language	0.13	0.39	0.328	0.743
Age	**0.11**	**0.04**	**2.940**	**0.003**
Pre/post: home language	**-0.57**	**0.22**	**-2.585**	**0.010**
Pre/ post: age	**0.05**	**0.02**	**2.315**	**0.021**
Home language: age	-0.03	0.04	-0.699	0.484
Pre/post: home language: age	**0.07**	**0.23**	**3.195**	**0.001**

Table 3. The GLMER model illustrates how age as well as, the effects of speaking one vs. two languages at home, influence correct responses in the oddity task at pre test and post test.

As in previous analyses, there was a significant main effect of pre/post-test, $\chi^2(1) = 44.02$, $p < 0.001$, indicating a training effect. A main effect of age was also observed, $\chi^2(1) = 15.50$, $p < 0.001$, with older children outperforming younger children. Average accuracy rates were 66.4% for 6–8-year-olds (n = 12), 78.2% for 8–10-year-olds (n = 18), and 79.4% for 10–12-year-olds (n = 25). No main effect was found for the number of home languages, $\chi^2(1) = 1.48$, $p = 0.224$, but a significant interaction between pre/post-test and number of home languages, $\chi^2(1) = 7.99$, $p = 0.005$, revealed that children who spoke two languages at home showed greater improvement (6.7%) compared to those who spoke one language (3.4%). A three-way interaction between number of home languages, pre/post-test, and age, $\chi^2(1) = 10.21$, $p = 0.001$, further clarified these effects (illustrated in Figure 2.6.). Among 6–8-year-olds, children who spoke one language at home improved more (5.1%) than those who spoke two languages (-4.2%). To assess this difference, we fitted the same model to responses from this age group only. The sub-analysis revealed a significant effect of language background for this age range, $\chi^2(1) = 4.55$, $p = 0.033$. Among 8–10-year-olds, improvements were similar for children with one home language (3.9%) and two home languages (6.7%), with no significant difference, $\chi^2(1) = 2.86$, $p = 0.091$. Among 10–12-year-olds, children who spoke two languages at home improved significantly more (9.8%) than those with one language (2.2%), $\chi^2(1) = 14.33$, $p < 0.001$. In summary, the best learners were older children who spoke two languages at home, with a bigger age effect among bilingual children.

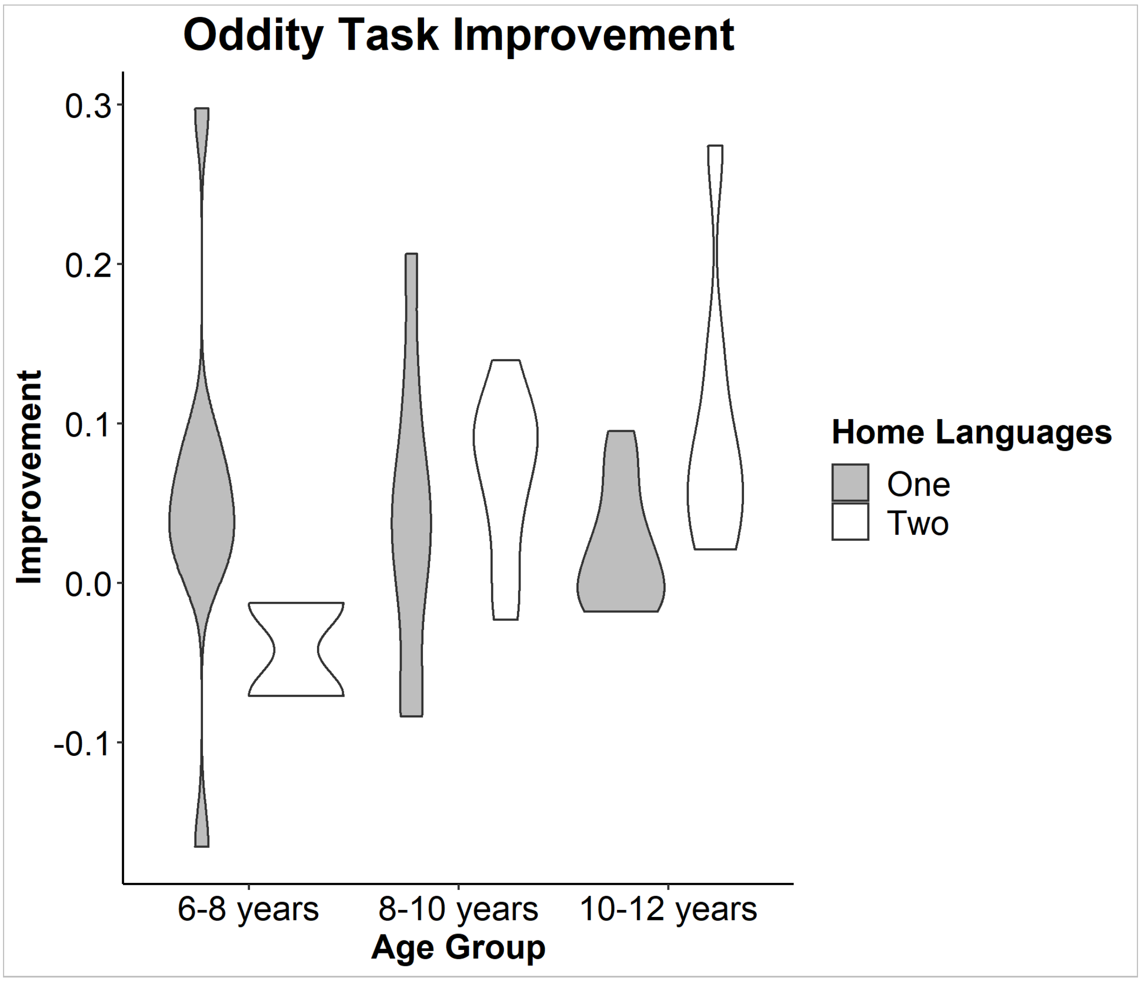

Figure 2.6. Violin plots illustrating the range of improvement in category discrimination accuracy between pre- and post-tests for children aged 6–8, 8–10, and 10–12 years from single-language and two-language homes. The greatest improvement in category discrimination was observed among older children from two-language homes.

To explore learning in the memory-card game, a mixed-effects model was used with the log-transformed time taken to complete each game as the dependent variable (see **Table 4**). The model included a linear fixed effect of session (training sessions 1, 2, or 3), a categorical fixed effect for symbol type to differentiate between the easy-symbol, hard-symbol, and no-symbols conditions, a linear fixed effect of age, and a categorical fixed effect for language background (one or two languages spoken at home). Additionally, by-subject random intercepts were included.

A significant training effect was found, $\chi^2(1) = 33.76$, $p < 0.001$, with the time taken to complete each game decreasing progressively from session 1 (74 s) to session 3 (66 s); see also Figure 2.7. A significant main effect of age was observed, $\chi^2(1) = 33.76$, $p < 0.001$, with 6–8-year-olds taking longer per game on average (106 s) than 8–10-year-olds (68 s) and 10–12-year-olds (55 s). However, no significant interactions between age and home language were found, unlike those observed in the pre/post tests, $p > 0.05$. Nevertheless, the time taken to complete the memory-card game is a less sensitive measure of phonetic perception, as it is influenced by multiple factors (e.g., memory, random card positions), compared to the pre/post categorical discrimination test.

	Time		
Fixed Effects	Estimate	SE	t value
(Intercept)	5.78	0.20	28.94
Easy symbols	-0.62	0.26	-2.35
Hard symbols	-0.01	0.27	-0.05
Session	-0.16	0.07	-2.21
Home language(s)	-0.64	0.20	-3.22
Age	-0.13	0.02	-6.55
Easy symbols: age	0.01	0.03	0.29
Hard symbols: age	-0.03	0.03	-1.06
Session: age	0.01	0.01	-1.06
Home language(s):age	0.07	0.02	3.41

Table 4. The LMER model illustrates the influence of the symbols condition and the age of the children on the time taken to complete a game in the memory card game, in each training session. The model also accounts for natural variability between participants and within experimental sessions.

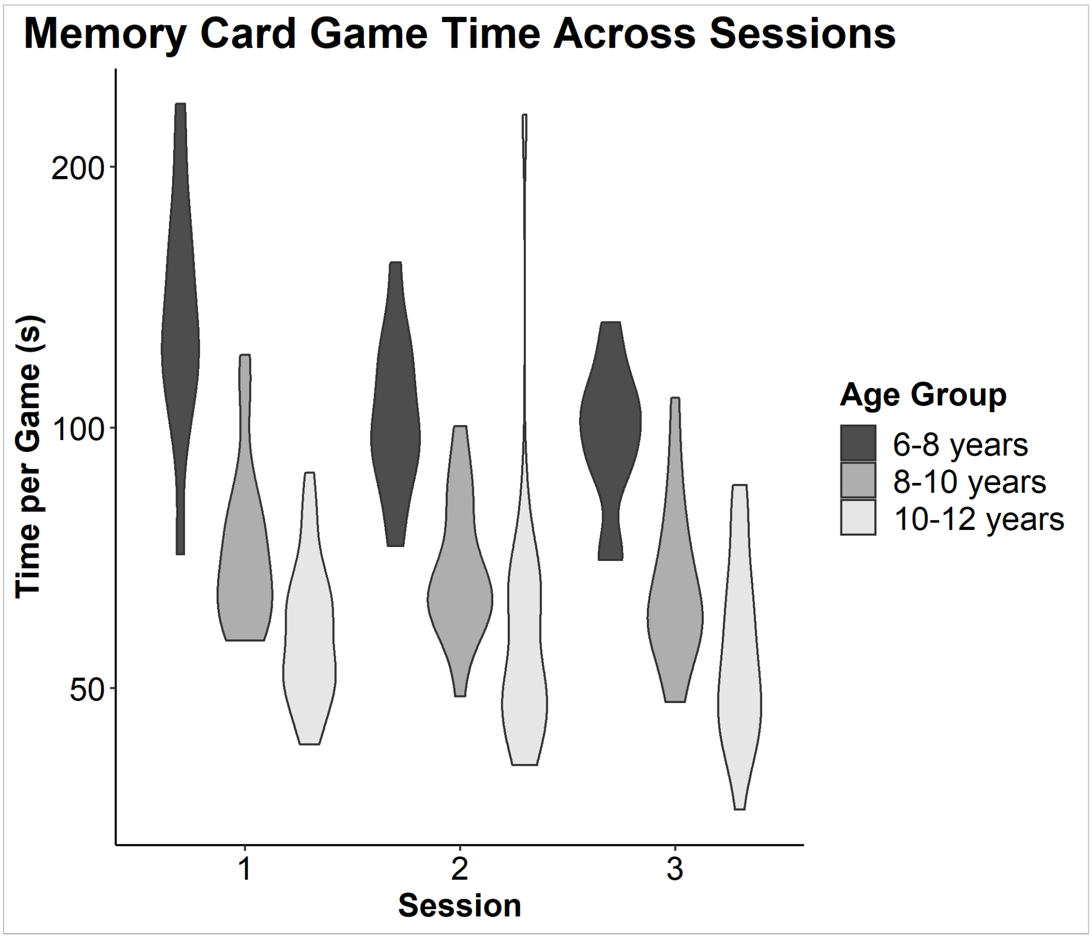

Figure 2.7. Violin plots showing the range of time taken to complete a memory card game during each training session for children aged 6–8, 8–10, and 10–12 years. Participants became faster at the game over the course of training, and the time taken to complete a game decreased progressively with age.

The memory card game was influenced by the type of symbols on the cards, $\chi^2(1) = 962.10$, $p < 0.001$, with participants taking longer, on average, in the no-symbols condition (74 s) compared to the hard-symbol (59 s) and easy-symbol (48 s) conditions; see Figure 2.8. An interaction between symbol type and age was observed, with the age effect being smaller for the easy-symbol condition (averages of 71, 51, and 40 s) compared to the hard-symbol (averages of 108, 64, and 51 s) or no-symbols (averages of 130, 81, and 67 s) conditions. However, no significant interaction was found between symbol type and pre/post-test or other factors, $p > 0.05$, suggesting that participants improved similarly across symbol types, despite the auditory-only condition being the most challenging.

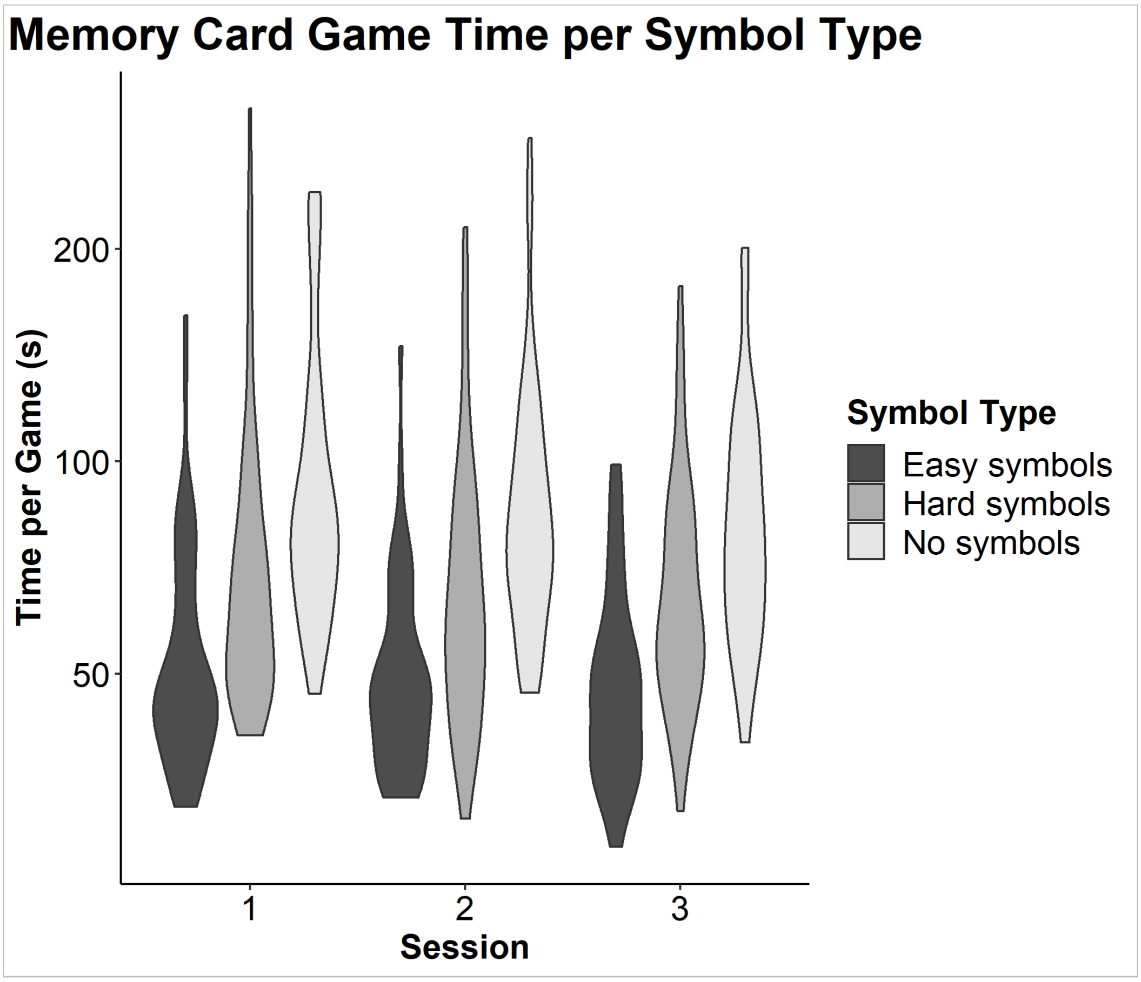

Figure 2.8. Violin plots illustrating the range of time taken to complete a memory card game in each training session, categorized by symbol type. Participants completed games the fastest in the easy-symbol condition and the slowest in the no-symbols condition. The relative speed of game completion for each symbol type remained consistent across all three training sessions.

2.4. Discussion

The current study used a phonetic training game that young children enjoy playing and in which they tend to perform well. The results still showed better performance in the older children like in previous child training experiments (Brekelmans, 2020; Kasisopa et al., 2018; Shinohara and Iverson, 2013, 2021). The training effect may seem small and that is because children were trained on fourteen vowel contrasts with varying degrees of difficulty. It is well known that Spanish and Catalan speakers struggle with the SBE contrast of /i/-/I/, as neither the Spanish nor Catalan repertoires have this tense-lax distinction, as well as

/æ/-/ʌ/, since Spanish and Catalan have only one vowel in that acoustic region (i.e., the low central vowel /a/) (Cebrian & Carlet, 2014). Moreover, the game included pairs that were anticipated to be easy for all children during the pre-test (e.g., /i/-/u/), and an improvement in discriminating these might have been difficult. Due to the practicalities of classroom testing, children played significantly fewer games than adults would play (Iverson et al., 2023). Despite these constraints, we consistently observed that older children showed greater learning. Additionally, this effect was influenced by home language, with older children from bilingual homes learning the most.

So why were 6-year-olds the worst performers? Perhaps, non-native phonetic learning is slow at that age. Burnham (2003) proposed that perceptual attunement is not uniquely confined to infancy but continues into early childhood and becomes especially important at the age of 6 years, as reading instruction begins. This entails a refinement in children's perceptual systems whereby they fine-tune their sensitivity to subtle phonetic distinctions within their native language to aid their phoneme-to-grapheme mapping. Increased attention to native phonemes leads to a diminished sensitivity for non-native contrasts and this could underlie the pattern of performance seen in the oddity task. Moreover, this reduced sensitivity for non-native contrasts at 6 years of age could be heightened in bilinguals who are trying to learn the phoneme-to-grapheme mapping in two languages. Having a wider native phonetic repertoire may thus, put bilingual 6-year olds at a disadvantage when acquiring the sound system of an additional language. Despite indications from the current data, further investigation is required to definitively attribute these observed outcomes to reduced sensitivity to L2 phonemes in 6-year-olds. We did not conduct additional assessments of language proficiency to discern whether this reduced sensitivity is causative or correlated. Nonetheless, future inquiries could extend to younger cohorts, such as four-year-olds, to ascertain whether a U-shaped pattern of performance emerges, similarly to Burnham's (2003) findings of children's sensitivity to non-native phonemes.

It is important to recognize that computer-based phonetic training may differ significantly from real-world language learning. Early research on high-variability phonetic training focused on naturalistic variability (e.g., Lively et al., 1993, 1994; Logan et al., 1991), making the training resemble an intensified form of real-world exposure. However, subsequent studies have shown that native French speakers living in London, who regularly use English, still benefit from phonetic training despite their extensive exposure to natural English speech (Iverson et al., 2012). This suggests that the structured listening environment of training allows for focused attention on phonetic contrasts in a way that everyday exposure might not. Additionally, training protocols may not need include stimuli with naturalistic variability (Brekelmans, 2020; Iverson et al., 2005; Sadakata & McQueen, 2014). Research indicates that, for adults, training often enhances the application of pre-existing phonetic categories rather than reshaping them to match the specific acoustics of the training materials (Iverson et al., 2005; Iverson & Evans, 2009). For instance, German speakers learn English vowels more quickly than Spanish speakers, likely because they can leverage their larger inventory of native-language vowels (Iverson & Evans, 2009).

Bilingual language use has a direct impact on the phonetic repertoires of bilingual speakers. Children who spoke two languages at home tended to use both of their languages more frequently. In adulthood, heightened co-activation of languages often results in greater phonetic convergence between the two languages (Simonet & Amengual, 2020). This means that increased use of both languages leads to greater similarity in the phonemes of each language. During childhood, however, phonemic categories are still developing, and this phonetic convergence may instead manifest as enhanced flexibility in forming vowel categories, which proves beneficial in non-native phoneme learning. These findings align with the observation that the most successful learners in our study were older children who are exposed to multiple languages at home.

Phonetic training may benefit from the increased metalinguistic awareness and enhanced phonetic and phonological knowledge that are found in bilinguals and older children. Bilingual children tend to have a better understanding of the rules and patterns that make up language and are better able to reflect on that knowledge (Jensen, 2008). Children from bilingual homes may inherently have a deeper awareness, that acquiring a new language entails learning a distinct set of phonemes. Consequently, they might demonstrate a better ability to identify non-native sounds. Additionally, older children in our sample have more experience with English lessons and may have acquired improved strategies to learn the language. As a result, older children from bilingual homes may be especially aware of what cues to look for in the training stimuli. On top of this, the more time children engage with their native languages, the more experience they gain with their native phonetic repertoires, consolidating their knowledge of a wider range of vowel contrasts. This broader phonological knowledge, could thus, contribute to the advantage seen in older children from bilingual homes in L2 phonetic learning.

Considering the role of attention in phonetic training, it is important to recognize the different attentional tendencies between children and adults. Children are less likely to maintain focus on tasks involving unfamiliar and repetitive sounds, unlike adults who are often driven by the long-term benefits, such as improved proficiency in a second language. In our 'easy symbols' condition, children's attention was captured by playful icons like emojis, which they matched to complete games. This approach is similar to tasks used by Lim and Holt (2011) and Saito et al. (2022) in their adult studies, providing a way to assess implicit phonetic learning in children. Our study found similar performance across all conditions, with no indication that one condition led to better learning. While Lim and Holt (2011) and Saito et al. (2022) saw phonetic learning during training, this effect was not significant from pre- to post-test, suggesting that focused attention on phonemes aids learning. However, our results indicate that children still show phonetic learning even when their attention is divided between auditory and visual-spatial tasks.

Altogether, the current study has demonstrated that the memory-card phonetic trainer helps enhance Spanish-Catalan children's perception of the Standard British English (SBE) vowel set. Notably, older bilingual children appear to benefit the most from the trainer, potentially because of their heightened metalinguistic awareness and expanded phonetic repertoires, resulting from exposure to multiple languages. While the poorer performance observed in 6-year-olds could be due to a potential second period of perceptual attunement, further research involving younger populations and further language assessments is needed to reach definitive conclusions. Additionally, future studies exploring the integration of visuomotor skills during training may offer insights into facilitating children's learning of non-native phonemes. Overall, our findings show that our platform can be used to study phonetic learning across the lifespan and suggest avenues for future investigation into the mechanisms underlying language acquisition in diverse linguistic contexts.

3. Mapping perception of the vowel space with Mismatch Negativity (MMN) response analysis.

3.1. Introduction

The process of perceptual attunement in infancy and the formation of native phoneme categories throughout childhood and adolescence, interact with phonetic learning in adult life. This interaction between early phonetic learning and later speech perception can be examined in the laboratory using behavioural tasks of identification and discrimination of phonemes. However, these tasks tend to be long and it is difficult to know whether listeners' responses reflect automatic speech processing or conscious, subjective judgments made after phoneme recognition. To untangle the impact of these factors on phoneme identification and discrimination, we can use neural measures, such as the Mismatch Negativity (MMN) response of electroencephalography (EEG) and magnetoencephalography (MEG) recordings. The MMN is elicited when an infrequent deviant sound is perceived as different from the recurrent standard stimulus and larger perceived differences between the standard and deviant stimuli, elicit larger MMN magnitudes (Näätänen et al., 2007a). Therefore, the MMN response provides a more direct measure of the brain's capacity to distinguish between stimuli, unlike behavioural measures, which can also be affected by attention or motivation. As a result, MMNs have been widely used to explore acoustic processing in the brain and infant and child studies have shown that these younger populations exhibit MMNs similarly to adults (e.g. Partanen et al., 2013).

The aim of the current experiment is to develop a time-effective MMN tool for testing perception of 19 IPA vowels in a single study. This tool could be used with subjects of different age groups and language backgrounds to track perception of the vowel space and provide insights into phonetic learning throughout development and across different speaker groups. Since MMNs are good indicators of discrimination accuracy, early studies investigated whether MMNs could be used as an index of phonological processing. Specifically, they looked at whether the MMN could be used to probe categorical perception in the brain (Aaltonen et al., 1997; Maiste et al., 1995; Sams et al., 1990; Sharma et al., 1993; Sharma & Dorman, 1998). However, in these initial studies, listeners' MMNs did not show increased magnitudes when the deviant stimuli crossed a category boundary, relative to the standard. Nonetheless, these studies measured perception of phoneme exemplars along a spectrum that systematically varied in a single acoustic dimension hence, several confounding factors could have influenced their results. For example, deviants with certain acoustic properties, such as higher second formant (F2) frequencies elicit larger MMNs making it difficult to discern the relative impact of acoustic versus phonological factors to MMN generation.

A way to disentangle the acoustic and phonetic/phonological contributions to the MMN is by testing phoneme perception across languages. Several cross-language studies have provided evidence that MMNs are sensitive to pre-attentive patterns of phoneme categorisation. For example, (Näätänen et al., 1997) played Finnish-speaking adults, sequences of standard /e/ vowels and three different deviant vowels: the Finnish vowels /ö/ and /o/; and the Estonian vowel /õ/. Although the acoustic difference between the Estonian vowel /õ/ and the standard /e/ is larger than that between the standard and /ö/, the MMN elicited by /õ/ had a smaller amplitude than those elicited by the Finnish vowels. In Estonian-speaking listeners, for whom all of these sounds were native vowels, the MMN magnitude was only dependant on acoustic deviation from the standard. Therefore, /õ/ elicited larger MMNs. This indicates that MMNs could be used to probe speech sound representations in the brain. Moreover, Cheour et al., (1998) showed that these memory traces for speech sounds emerge early on in development. They carried out a similar experiment with 6-month and 12-month-old Finnish and Estonian babies and saw that the

6-month-old Finns had larger MMNs for the Estonian /õ/ than the Finnish /ö/, i.e., the acoustically less deviant stimulus. At 12-months, these infants had a much larger MMN for the native /ö/ than for /õ/, suggesting that the language-specific memory traces developed between 6 and 12 months. The Estonian 12-month-old infants for whom both deviants were native vowels, had larger MMNs for the acoustically more different /õ/, like the Estonian adults. Therefore, this evidence suggests that MMNs can serve as a valuable tool for exploring phonological processing across various age groups.

Additionally, studies have found increased MMN sensitivity for native-like consonantal phonemes. Dehaene-Lambertz, (1997) showed that, in French listeners an MMN was elicited for the across-category distinction between /ba/ and /da/, while no MMN was seen for the dental versus retroflex distinction, which is not phonemic in French. Similarly, Sharma & Dorman (1999) examined Voice Onset Time (VOT) in English speakers and observed that an across-category change from 30 ms to 50 ms VOT elicited an MMN, while a within-category change from 60 ms to 80 ms VOT did not. In a later study, Sharma and Dorman (2000) used syllables starting with pre-voiced stop consonants with VOT values of −10 ms and −50 ms as standard and deviant stimuli. Native English listeners categorized both stimuli as /ba/, whereas Hindi listeners classified the −10 ms VOT as /pa/ and the −50 ms VOT as /ba/. Importantly, only Hindi listeners showed significant MMNs, highlighting the phonetic distinctions relevant to their language.

The reviewed studies show that there is a clear phonological contribution to MMN generation. There are several theoretical frameworks for the MMN, which aim to explain the perceptual process that it reflects, at different levels of auditory processing. The most prominent are the neural adaptation hypothesis and the predictive coding hypothesis. The neural adaptation hypothesis explains MMN generation at lower levels of auditory perception and describes a basic sensory mechanism for detecting changes in sound sequences.

According to the neural adaptation hypothesis, the standard stimuli causes a reduction of neural responses due to synaptic fatigue or habituation. When participants hear a deviant sound, a stronger response is elicited because the neurons are not adapted to it. Under this view, the MMN is seen as a result of differential adaptation to standards and deviants (Garrido et al., 2009; Jääskeläinen et al., 2004). The predictive coding hypothesis explains how higher-order expectations and context influence the MMN. The brain generates continuous predictions about incoming sensory input based on prior experience. When an incoming stimulus deviates from these predictions, it generates prediction errors, which are used to update the brain's internal models. In the predictive coding hypothesis, the MMN is taken to reflect the process of updating the brain's internal model (Garrido et al., 2008, 2009; Winkler, 2007).

The MMN's sensitivity to native phonemic categories, reflects, not only short-term memory processes, but also demonstrates effects of long-term memory traces representing native phonemes. That is, MMNs to phonemic deviants are larger than the MMN elicited purely to the acoustic deviance between the standard and the deviant stimuli. Moreover, neuroimaging studies have shown different patterns of activation in MMN responses to non-linguistic sound changes and language-specific MMNs. The former show stronger activation in the right hemisphere compared to the left, whilst language-specific MMN sources predominantly activate the left temporal lobe (see Näätänen et al., 2007). Therefore, analysing MMN magnitudes elicited by different phoneme deviants should provide valuable insights into individuals' pre-attentive perceptual spaces.

However, there are several problems with the traditional way of MMN testing. Firstly, the MMN is a small neural response and many trials are needed for the deviant stimuli to get reliable MMNs i.e., approximately 100 deviant stimuli per deviant type (see Kujala et al., 2007). Additionally, MMN paradigms require a small percentage of deviants, as these must be perceived as infrequent, with sufficient time gaps between the stimuli to accurately record neural responses (see Näätänen et al., 2007). In traditional event-related potential (ERP) analysis, data is typically taken from the channel with the best signal (often Fz, F3, or F4), and statistical tests are conducted on that single channel for each subject, requiring many EEG epochs to ensure reliability. Consequently, the time needed to test a single standard-deviant contrast is lengthy and as a result, most MMN investigations only include one type of deviant per oddball sequence.

Recent investigations have introduced new oddball paradigms designed to reduce the duration of MMN testing. For example, Näätänen et al., (2004) introduced the 'multifeature paradigm' in which every other stimulus is a standard and every other, is one of five deviants. This paradigm relies on the assumption that the deviants strengthen the memory trace of the standard by reiterating the features that both types of stimuli have in common. Partanen et al., (2013) showed that, even in infants, the magnitude of the MMNs elicited in a traditional oddball and the multifeature paradigms did not differ. Moreover, conveniently for infant experiments, researchers were able to explore the perception of five standard-deviant contrasts in a 40-minute testing session. Additionally, researchers in the auditory world have leveraged the fact that MMNs also reflect listeners' ability to passively identify abstract patterns in sound sequences by incorporating various types of deviants in their studies. For instance, Koelsch et al., (2016) presented participants with sequences of six timbres organized into triplets, forming a tonal "alphabet" of nine possible triplets. These triplets varied in rarity (10%, 30%, 60%), though each timbre's occurrence was consistent. Both rare and intermediate triplets elicited significant MMN, showcasing the brain's capacity to track complex regularities (Tsogli et al., 2019).

In addition, innovative MMN analysis techniques have been developed to shorten testing time. Brandmeyer et al., (2013) succeeded in using machine learning (ML) techniques to analyse single-trial MMN data. In their study, they used multivariate pattern classification methods to train a classifier which divided neural responses into two groups, depending on whether they followed a standard or deviant stimulus. The classifier's weights were adjusted depending on the spatiotemporal patterns of the recorded EEG data linked to each trial type. A ten-fold cross-validation analyses was used to train and test the algorithm. Meaning that, the EEG data from each subject was divided into ten sections and predictions were made for each section of data using the remaining nine segments. Their classifier was shown to reliably predict whether new incoming neural data was a response to a standard or deviant and analyzing MMN data at the single-trial level reduces significantly the length of testing sessions.

The current study aims to train a model which will predict MMN magnitude in a continuous manner using similar ML techniques to those described above. In the current investigation, the EEG data is analysed with backwards multivariate Temporal Response Functions (mTRFs; Crosse et al., 2016). This MATLAB toolbox has been widely used to study neural tracking of the speech envelope (e.g. Attaheri et al., 2022; Zinszer et al., 2022) and we hypothesise that mTRFs will be well-suited for MMN analyses. Unlike traditional ERP analysis, backwards model mTRFs use the whole EEG dataset to build predictive models of the stimulus input and therefore, gather greater statistical power. These models are built using regularised regressions of the EEG data and stimulus input functions, to prevent overfitting. They are then, trained using leave-one-out cross-validation, i.e., the neural response to each stimulus is predicted based on a model trained on all the other stimuli, omitting each stimulus from its own training set. Evaluation of the model involves correlating the predicted and actual stimulus functions, with high correlation indices reflecting high levels of prediction accuracy.

We use a novel oddball paradigm to elicit MMN responses. It consists of standard and deviant vowel pairs. Standard pairs are made up of two distinct vowels. The first vowel is kept constant for several trials and the second vowel changes randomly in every trial (e.g., a-e, a-u, a-i). Deviants consist of two repeating vowels (e.g., a-e, a-u, a-i, **a-a**). To make the oddball sequences, we downloaded 19 IPA vowel recordings from the Seeing Speech (2014) website. The selected vowels are represented with the following IPA symbols: /ɨ /, /ʉ /, /ø/, /œ/, /ɶ/, /ɒ /, /ʌ /, /ɤ /, /ɯ/, /i/, /e/, /ɛ /, /a/, /ɑ /, /ɔ /, /o/, /u/, /y/, /ə /. These recordings were made by a female phonetician and for each vowel, we processed the signal to give it natural, uniform amplitudes and pitch contours. One vowel pair was played every second: each vowel pair had a duration of 400ms; the second vowel started 200ms after the first one; and there were 600ms gaps between pairs. These timings were chosen to allow us to record changes in neural activity for each vowel pair and draw comparisons between standard and deviant pairs. In the oddball paradigm, 15% of tones were deviant vowel pairs, picked randomly, and each deviant was always followed by at least one standard. Figure 3.1. below shows a spectrogram of the vowel sequence. Standard trials are marked with an S and deviant trials, with a D.

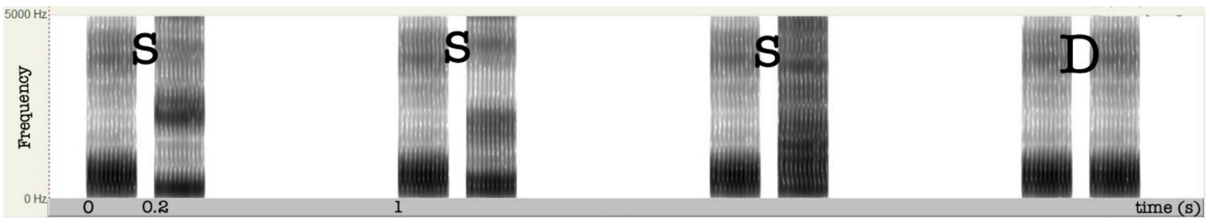

Figure 3.1. Spectrogram of a section of the oddball paradigm.

Our main aim is to develop a methodological framework to shorten the length of testing whilst still eliciting MMN responses to the deviants. For this purpose, a simulation is carried out using permutation statistics of our modelled EEG data. With this analysis we aim to explore whether a shorter version of the paradigm could still provide sufficient data to evaluate potential mismatch responses. To investigate this, we run a simulation to estimate how many deviant trials per deviant type would be necessary to detect reliable differences between standard and deviant responses. This approach involved resampling and permutation analyses of both ERP and mTRF data, allowing us to assess the robustness of possible mismatch effects across different trial counts. By identifying the minimum number of trials needed to detect a reliable signal, this analysis provides insight into how the paradigm might be optimised for future use.

Our secondary aim is to explore how reliable the recorded MMN magnitudes are at indexing vowel categorization. For this purpose, an online vowel identification (ID) and category goodness task is delivered to participants. The same set of 19 IPA vowels are used in the MMN experiment and the behavioral tasks. The vowels that are consistently assigned to a single native category, in the vowel ID task are assumed to be perceived as belonging to a native category. Therefore, repeated vowel pairs with high ID ratings are expected to elicit the greatest MMNs. The findings from this study will contribute towards our understanding of MMN generation and will be useful for collecting extensive data to assess the development of speech discrimination abilities across the lifespan.

3.2. Methods

3.2.1. Subjects

A total of 17 native British-English adults (M=28.42 years, SD= 9.60 years) participated in the EEG experiment. Data from 4 participants was excluded due to issues with the trigger box so 13 participants were included in our analyses. To assess whether the MMN magnitudes would index vowel categorization, a Vowel ID and Category Goodness task was delivered to 14 native British-English adults (M=30.38 years, SD= 9.28 years). Participants were students or staff members at University College London (UCL) and were contacted through the UCL Psychology Subject Pool.

3.2.2. Materials

The stimuli used in the oddball paradigm consisted of 19 distinct vowel recordings selected from the Seeing Speech (2014) database. These vowels were recorded by a female phonetician and represented a wide range of phonetic categories, with the following IPA symbols: /ɨ/, /ʉ/, /ø/, /œ/, /ɶ/, /ɒ/, /ʌ/, /ɤ/, /ɯ/, /i/, /e/, /ɛ/, /a/, /ɑ/, /ɔ/, /o/, /u/, /y/, and /ə/.

To ensure consistency and naturalness across all vowel sounds, the recordings were processed to standardize amplitude levels and pitch contours. This processing aimed to eliminate any variability in the recordings that could introduce confounds during the auditory presentation.

The vowels were paired in different combinations to form both standard and deviant vowel pairs, with the standard pairs consisting of two distinct vowels and the deviant pairs consisting of two identical vowels. These vowel pairs were then presented as auditory stimuli in the oddball task.

3.2.3. Procedure

Participants were invited to take part in an EEG recording in the laboratory. Testing sessions were carried out in a sound-proof EEG booth where participants took part in a passive listening task whilst they watched muted cartoons. In the EEG recording session, participants heard 12 oddball sequences of concatenated standard and deviant vowel pairs, each lasting around two and a half minutes. Breaks were allowed between the sequences and the overall EEG recording time was of 45 minutes.

In the oddball sequence, standard pairs comprised two different vowels: the first vowel remained constant across several trials, while the second vowel varied randomly (e.g., *a-e*, *a-u*, *a-i*). Deviant pairs, by contrast, consisted of two identical vowels (e.g., *a-a*), and were interspersed randomly among standard trials with 15% of the trials being deviant vowel pairs, selected at random. Each trial consisted of one vowel pair presented over the course of one second. Within each pair, the first vowel had a duration of 400 ms, and the second vowel began 200 ms after the onset of the first. A 600 ms inter-trial interval followed each pair. To maintain the structure of the paradigm, each deviant was followed by at least one standard trial. Figure 3.1. provides a spectrogram of a sample stimulus sequence, with standard trials labeled 'S' and deviant trials labeled 'D'. A total of 1920 vowel pairs through Etymotic ER-1 insert earphones.

3.2.4. EEG data recording and pre-processing

EEG was recorded through a Biosemi Active Two system with 64 (Ag/AgCl) electrodes mounted on an elastic cap and 6 external electrodes: left and right mastoids and two vertical and horizontal EOG electrodes. Recordings were made with a sampling rate of 2048 Hz and downsampled to 64Hz.

All EEG recordings were preprocessed using a custom MATLAB pipeline combining EEGLAB (Delorme & Makeig, 2004) and FieldTrip (Oostenveld et al., 2011) functions, along with additional routines from the NoiseTools (de Cheveigné & Simon, 2008) toolbox. For each participant's file, data were imported with a 64-channel montage and 8 external channels, initially referenced to Fz. Eye channels were used to compute vertical and horizontal electrooculogram (EOG) signals, which were filtered to isolate blink and saccade-related activity. The data was high-pass filtered with a zero-phase first-order Butterworth filter at 0.1 Hz and low-pass filtered with a zero-phase first-order Butterworth filter at 8 Hz, separately in each recording block, to remove slow drifts. Within each recording block, data were robustly re-referenced using a multichannel average reference approach that downweighted outliers, as implemented in the `nt_rereference` function from the NoiseTools toolbox. Windowing was applied at segment boundaries to reduce edge artifacts. Blink and horizontal eye movement artifacts were then attenuated using a Denoising Source Separation (DSS)-based spatial filtering approach, identifying and removing components that correlated strongly with vertical EOG and horizontal EOG activity.

3.2.5. Vowel Identification and Category Goodness Task

The vowel identification and category goodness task was created and administered using the Gorilla Experiment Builder so that participants could complete it remotely, using their phone, PC or laptop. The same 19 IPA vowels were used here as in the oddball paradigm. On each trial of the task, participants were prompted to press a button to play a vowel, listen to it and select which of the six words on the screen contained the vowel that they heard. The six response options linked to each IPA vowel were selected from a list of bVt minimal pairs (i.e., bait, Bart, bat, beat, Bert, bet, bit, bite, boat, boot, bot, bought, bout, but). Three female recordings of these minimal pairs were used to extract an average F1 for each word. Each IPA vowel was then linked to the six words that had the closest F1 averages. For the identification judgement, participants could only listen to the vowel once in each trial. Afterwards, they were prompted to judge the goodness of the vowel sound by positioning a pointer along a spectrum of "very good" to "very poor" exemplar. Each vowel was presented four times. Meaning that there were 76 trials in total, which were divided into four equal blocks to allow for breaks between blocks.

3.3. Results

The analysis of the EEG recordings and the vowel identification task were carried out separately and were later compared to evaluate the correspondence between the MMN magnitudes and the identification scores for each IPA vowel.

3.3.1. ERP analysis

In the first instance, we analysed whether our vowel-pair oddball paradigm elicited MMNs responses. For this purpose, we carried out a traditional ERP approach to observe whether there was a differential neural response to standard and deviant pairs.

For each subject, we selected the EEG channel with the best data and carried out an average ERP for standard and for deviant vowel pairs across all subjects. These averages show that our oddball paradigm elicited MMNs with peaks between 100-250ms after onset of the second vowel. Figure 3.2. shows a visual representation of these averaged ERPs in the time window starting 200ms before onset of the first vowel and ending 800ms after. The blue line represents responses to standard stimuli and the red line represents responses to the deviants. On the right-hand side, the ERP sensor space revealed left-lateralization of the MMN responses.

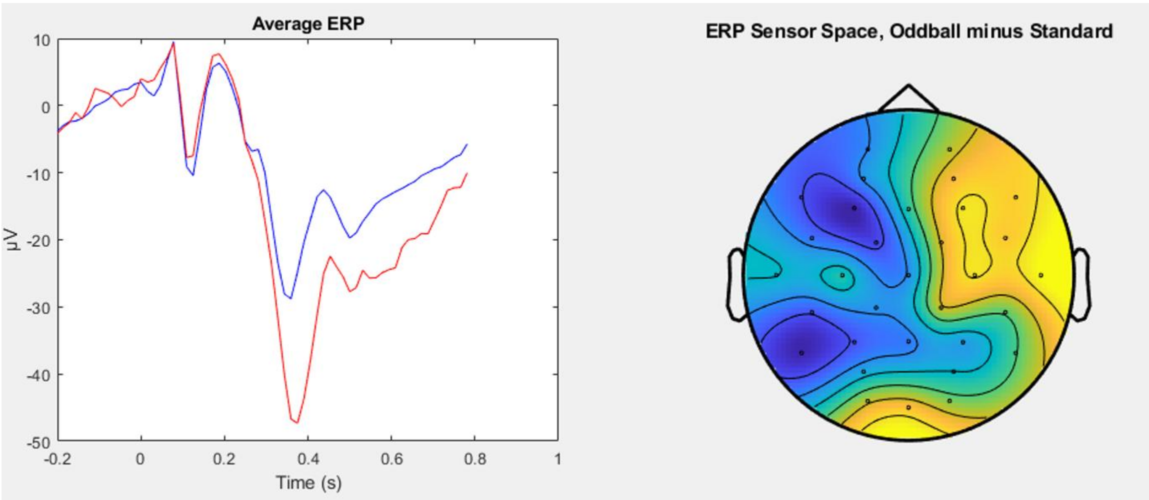

Figure 3.2. Plots showing the averaged event-related potentials (ERPs) and the sensor space. On the left, average event-related potential responses for all subjects. The blue line represents neural responses to standards and the red line, the responses to deviants. On the right, a scalp topography of the average MMN calculated by subtracting the standard responses from the deviants.

On the average ERP graph in Figure 3.2., there is a clear differential response between the standard responses (in blue) and the deviant responses (in red). For both response types, there is a P1, N1 and P2 upon onset of the first vowel in the pair. At 0.2s, the second vowel is played and there is a slight P1 in response to the second vowel. At time point 0.38s, i.e., around 160ms after onset of second vowel, there is a clear difference between the magnitude of the standard and deviant negative peaks –we get an MMN response. Here, the latency of the MMN peaks goes in line with those reported in the MMN review by Näätänen et al., (2007). Regarding the source of the MMN, our results show left-lateralized MMN sources which differ from the fronto-temporal sources reported by Näätänen et al., (2007). A reason for this observation could be that our deviants did not entail an acoustic difference like in most studies included in the review by Näätänen et al., (2007), resulting in a different activation pattern.

3.3.2. mTRF analysis

In the second step of the analysis we fit Multivariate Temporal Response Functions (mTRFs; Crosse et al., 2016) to our EEG data. This analysis aimed to assess whether mTRFs could serve as a suitable and time-efficient method for examining the MMN. Specifically, we aimed to investigate whether mTRFs could reliably capture the MMN response even when fewer deviant trials were included in the paradigm. Firstly we fit mTRFs to our data and visualized the mTRF model's predictions in several plots. Secondly we evaluated the statistical reliability of the mTRF analysis via a bootstrapping analysis.

MTRFs have mostly been used to study cortical tracking of the continuous speech envelope but they have also been used to analyze ERP components such as the N400 (Broderick et al., 2021). In the mTRF analysis, a stimulus function is created, with peaks upon stimulus onset and the raw EEG data is mapped back onto the stimulus function to reconstruct the speech signal that subjects listened to during the experiment.

To detect differences in neural responses between standard and deviant stimuli, we constructed a stimulus function that assigned differential weights to each trial type. Standard stimuli were given positive weights, while deviant stimuli were assigned larger-magnitude negative weights. These opposing and asymmetric weights served two purposes: (1) to model the expected contrast between standard and deviant responses, and (2) to account for the imbalance in trial counts due to the oddball design, in which standard stimuli occurred more frequently. By weighting deviants more heavily, we ensured that both trial types contributed equally to the analysis despite the frequency imbalance. This approach increased sensitivity to neural differences between conditions. Figure 3.3. below shows the stimulus function for standard pairs in blue and for deviants, in red.

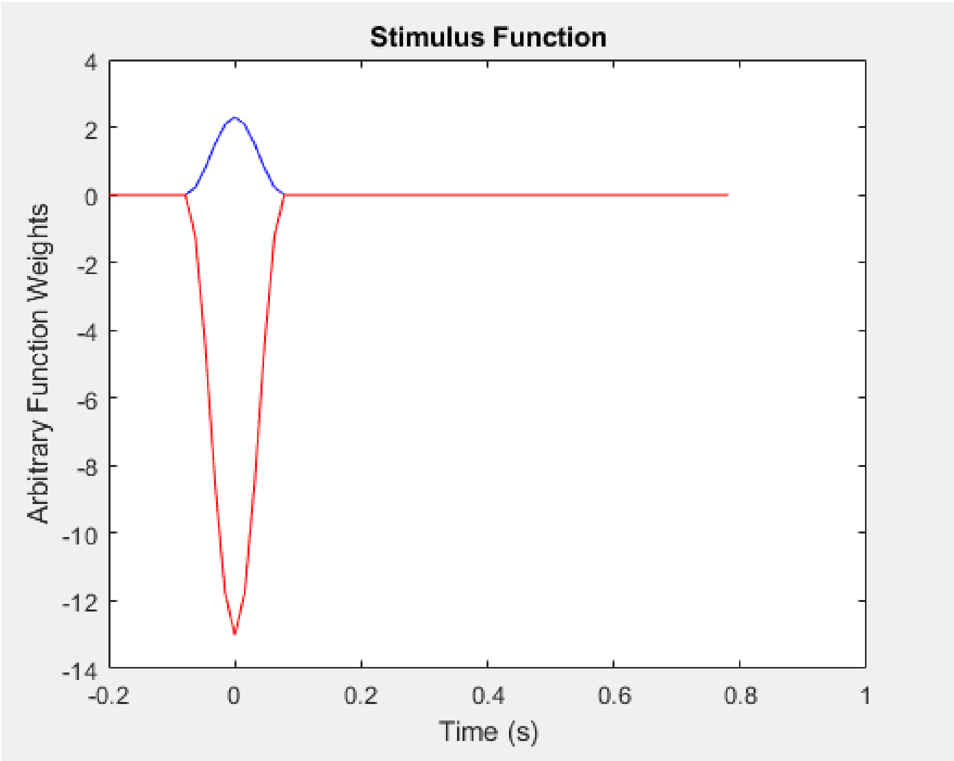

Figure 3.3. Stimulus function for standard and deviant stimuli. The stimulus function for the standards is represented in blue and the stimulus function for the deviants, in red.

We fit multivariate Temporal Response Functions (mTRF; Crosse et al., 2016) in backward models to relate the EEG activity of each subject to the stimulus functions. In this approach, weights were estimated for each of the 64 EEG channels across a range of time lags, allowing us to reconstruct the stimulus from the neural response. These weights act like a temporal filter, selecting the most important channels for stimulus reconstruction at each point in time. Our mTRFs were trained over a one-second window because a vowel pair played every second in the oddball paradigm and the key mathematical steps involved in calculating these weights is explained below.

The mTRF framework does not discard data from any electrode, therefore, the data going into the analysis is intercorrelated. To avoid overfitting, the mTRF approach carries out regularized regressions of the EEG data and the stimulus functions. Meaning that, it aims to minimize the sum of squared residuals plus a "penalty value", (i.e., *lambda*) between the predicted and actual stimulus values. The magnitude of lambda is indicative of the importance of the predictor at each point in time in the regression, with a high value of lambda implying that the predictor is less useful for stimulus reconstruction.

To calculate the most appropriate lambda values, a ten-fold cross-validation approach is used. Here, the EEG data for each subject is divided into 10 sections and the stimulus function corresponding to each segment of the EEG data is predicted based on a model trained on the remaining nine sections, omitting each EEG data section from its own training set. The value of *lambda* is adjusted to build the most accurate predictions that will generalize well to new incoming data. Once all these processes are carried, data from each subject yields a unique set of weights, (i.e., a 64x64 matrix) which is convolved with the EEG data to generate a predicted stimulus function.

We averaged the predicted stimulus functions across subjects to assess whether the resulting average resembled the actual stimulus function. Figure 3.4. shows the average stimulus function, the lambda coefficients and the mTRF sensor space. In the average mTRF graph, the blue line represents the averaged mTRFs for standard pairs and the red line represents the mTRFs for the deviants.

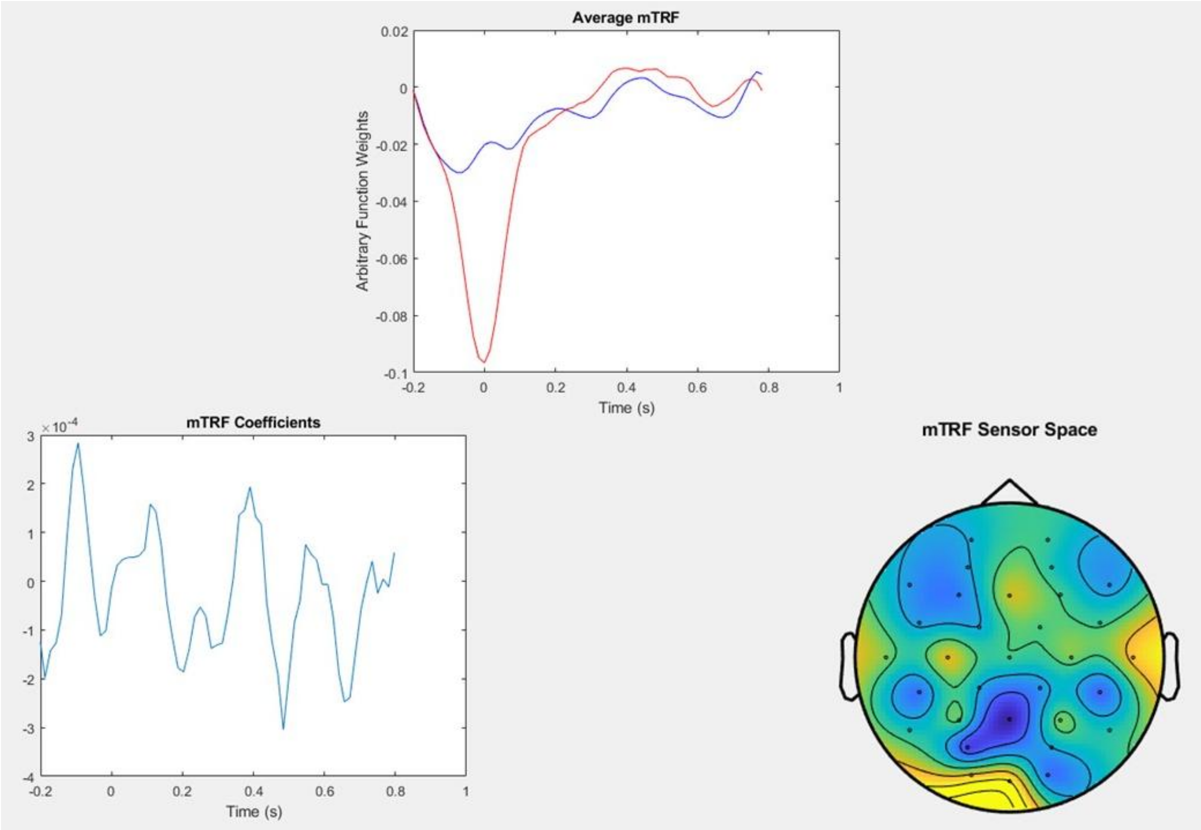

Figure 3.4. Plots showing the averaged mTFR results. Above is the averaged reconstructed stimulus function for all subjects with standards in blue and deviants in red. On the bottom left are the lambda mTRF coefficients and to the right, the mTRF sensor space.

In Figure 3.4., the averaged reconstructed stimulus function for the deviant stimuli closely resembles the input stimulus function used in the analysis. Specifically, it shows a large negative peak at time 0 s, matching the input function. The amplitude of the function weights is smaller in the reconstructed signal than in the original stimulus function, which is expected given the inherently low correlations typically observed in mTRF analyses. In contrast, the reconstructed stimulus function for the standard stimuli displays a different profile: it shows a small positive peak around time 0 s and exhibits similar function weights to the deviant reconstruction at approximately 100 ms after stimulus onset. This parallels the structure of the original input function. These results demonstrate that the neural responses to standard and deviant pairs yield distinct stimulus reconstructions, indicating a clear difference in EEG activity between the two conditions.

Regarding the mTRF coefficients, the peaks in the graph show the time points at which we obtained the most valuable EEG data for stimulus reconstruction. The mTRF sensor space aims to show a localization of the source of the MMN. Nevertheless, mTRFs also show data cleaning and noise reduction processes rather than purely source localization. The blue areas in the scalp topography represent the location of the electrodes that were most relevant for stimulus reconstruction in the one-second time window of analysis.

3.3.3. Simulation and significance testing

Once we established that the mTRF method allowed us to effectively reconstruct the stimulus function and observe differences in neural responses to standards and deviants, we aimed to evaluate the feasibility of shortening the oddball paradigm. Therefore, we carried out a simulation using permutation analyses of the ERP and mTRF data. In particular, we explored whether we would get reliable MMNs if our paradigm included fewer deviants per deviant type (N=19).

In the permutation analyses of our ERP and mTRF data, we included 7 standard responses for each deviant response. This allowed us to estimate the number of deviant trials required per deviant type to reliably detect MMN responses above baseline levels. To create a baseline measure, the data was rearranged and randomly split into standard and deviant responses using resampling with replacement. Significance tests were carried out to observe the proportion of times in which the actual MMN responses were greater than the baseline. Results revealed that, in our ERP analysis, we would need around 100 deviants per deviant type to get MMNs that were greater than the baseline, 95% of the time. This goes in line with previous findings (Näätänen et al., 2007b). With mTRFs, we were able to reach that level of reliability after 1 deviant per type of deviant. Figure 3.5. shows the probability thresholds for the ERP and mTRF data after *n* number of deviants per deviant type were included in our paradigm.

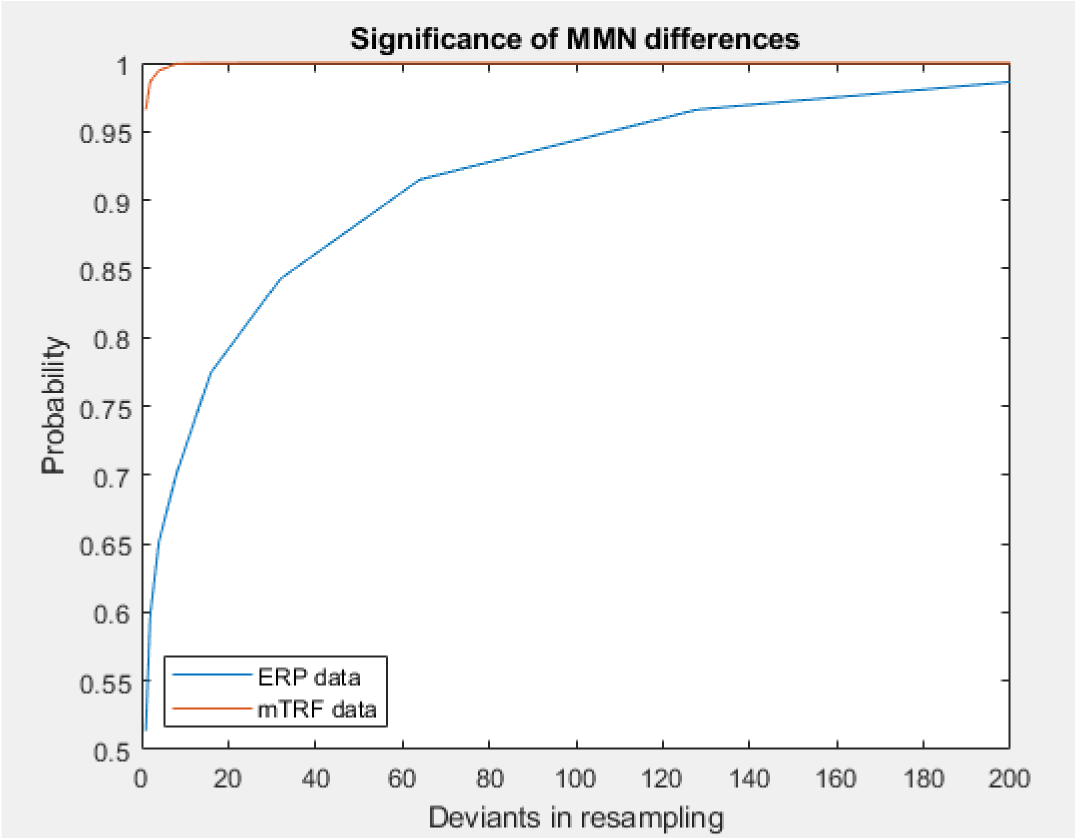

Figure 3.5. Plot demonstrating the bootstrapping results of significance testing of recorded MMNs. The graph shows the probability that the MMN recorded is not due to chance according to the number of deviants included in the paradigm per deviant type. Blue represents ERP data and orange, the mTRF data.

These results show that when MMN responses are analysed using mTRFs, less deviants are needed, per deviant type, in the oddball paradigm. Therefore, our methodological framework allows us to build oddball paradigms with multiple phonetic contrasts and shorten the length of MMN testing. It is important to consider that although our permutation analyses show reliable MMNs after one deviant per deviant type, the mTRF weights were calculated using a larger dataset. The key in future investigations would be to find a balance between reducing the number of standard and deviant stimuli used and collecting enough EEG data to build reliable predictive models.

3.3.4. Magnitude of MMN as an index of vowel categorization

A secondary aim of this investigation was to use the average MMN magnitude elicited by each repeated vowel deviant as an index of vowel categorization. We first explored whether the magnitude of the recorded MMN responses would go in line with responses on the vowel ID task. We calculated each IPA vowel's identification score as the proportion of times it was matched with its most frequently chosen word.The vowels with the largest ID score were taken to demonstrate the greatest assimilation into a native vowel category. Visual inspection of the data revealed no correlation between ID scores and MMN magnitudes and as a result, we also investigated whether there were acoustic driving factors of MMN magnitude. We extracted four spectral measures for each IPA vowel: the first (SM1) and second (SM2) spectral moments, which reflect the general position of energy in the spectrum and the spread of spectral energy, respectively, and the first (F1) and second (F2) formant frequencies, which correspond to resonant frequencies of the vocal tract associated with vowel height and backness.

To test the statistical interaction between the vowels' ID scores and acoustic parameters and the generated MMN magnitudes, we ran a mixed model analysis using the lme4 package (Bates et al., 2015) in R. We used the CAR package (Fox & Weisberg, 2011) to obtain p-values within a type II analysis of deviance table. The model had five fixed effects: ID scores, SM1, SM2, F1 and F2. Moreover, we included by-subject random intercepts. The results from the model demonstrated no significant main effect of ID scores $\chi^2(1)=.0019$, $p=.97$, indicating that ID scores were not a suitable predictor of MMN magnitude. Nevertheless, there was a significant main effect of SM2 $\chi^2(1)=7.81$, $p=.005$ and a main effect of F2 $\chi^2(1)=4.2$, $p=.04$, indicating that the MMN magnitudes were driven by acoustic factors rather than perceptual effects. The ID scores and MMN magnitudes of each IPA vowel are detailed in Figures 3.6. and 3.7., and the main effects of SM2 and F2 are plotted in Figure 3.8.

Firstly, we looked plotted each IPA vowel in an F1/F2 plot. In Figure 3.6. the size of the dot representing each IPA vowel, corresponds to the magnitude of the ID score, with larger dots demonstrating greater ID scores. Additionally, the word selected most often for each vowel is displayed underneath.

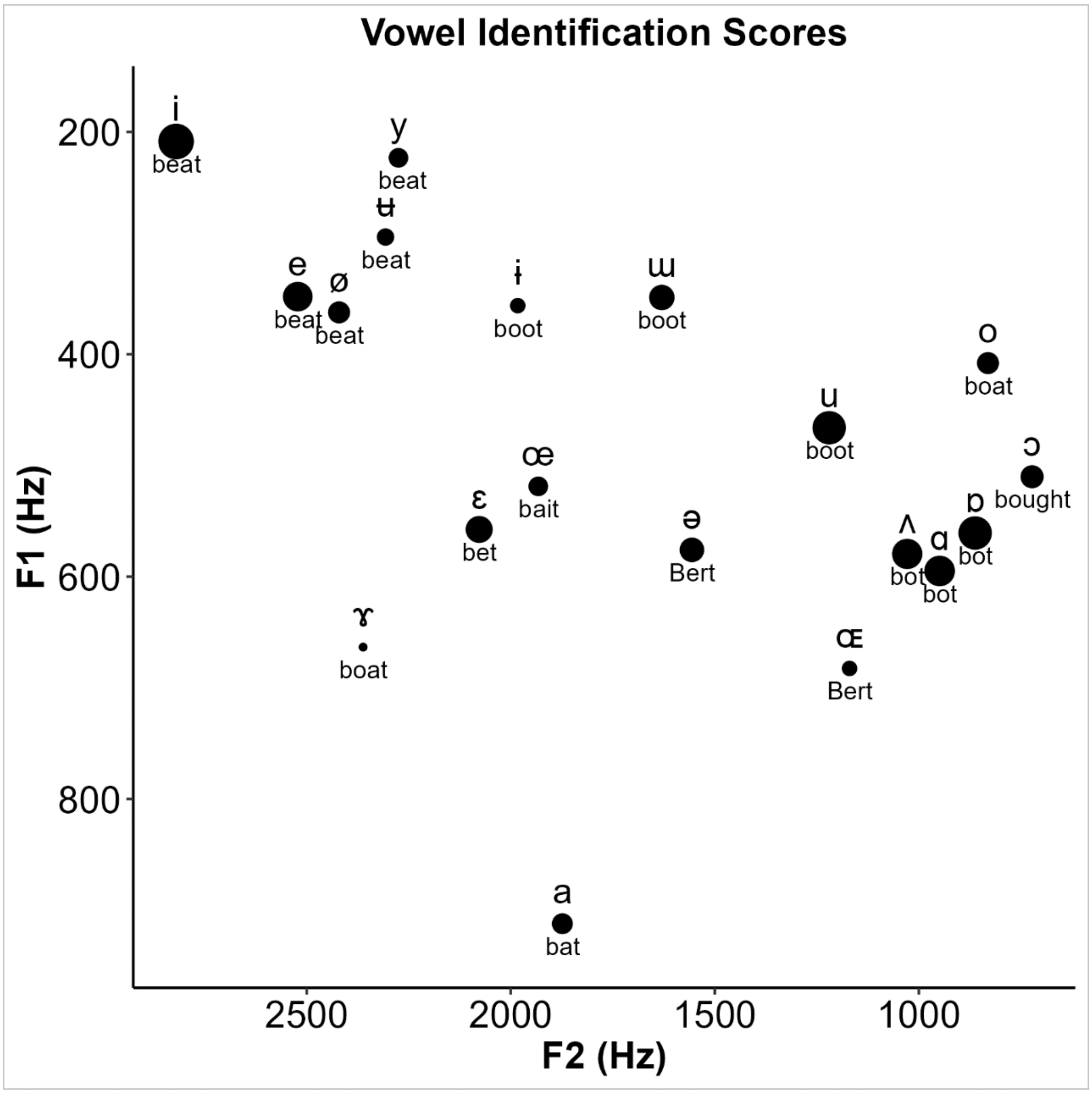

Figure 3.6. Formant plot illustrating the vowel identification scores. The plot shows the 19 IPA vowels plotted according to their F1 and F2. The size of the dot for each vowel represents the proportion of times that each vowel was mapped onto the word displayed underneath it. The bigger the dot, the greater the ID score.

Figure 3.6., demonstrates that high, front vowels were mostly perceived as the native /i/ and low back vowels, as the native /ɒ /. There were some surprising results. For example, /e/ was mostly linked to "beat" and not "bet" and /ʌ / was perceived as the vowel in "bot" and not "but". Therefore, our stimuli may not be good representations of native Southern British English vowels. Those vowels that were most consistently mapped onto a single

native category were /i/, /ɒ /, /u/, /ɑ / and /ʌ / and although, we initially predicted that deviant pairs of these vowels would elicit the largest MMNs, statistical results demonstrated that ID scores were not good predictors of MMN magnitudes. Figure 3.7. shows the same F1/F2 plot of the IPA vowels with the size of the dot representing the average magnitude of the MMN elicited across all subjects for each repeated vowel deviant.

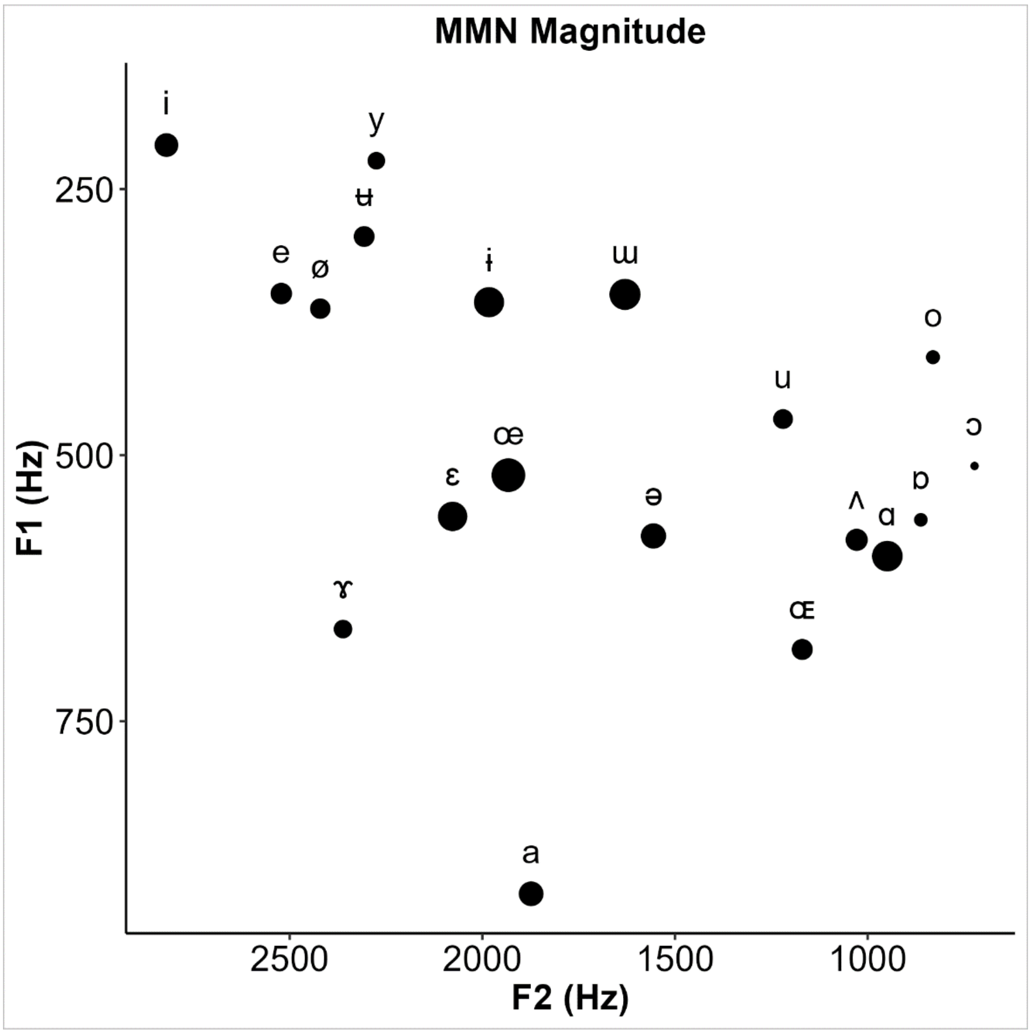

Figure 3.7. Formant plot illustrating the MMN magnitudes generated by each vowel. MMN magnitude is represented by the relative size of the black dot belonging to each vowel. The vowel deviants which yielded greater MMN magnitudes were /ɨ/, /œ/, and /ɯ/, which are not native vowels to our participants.

Lastly, Figure 3.8., shows scatterplots of the second spectral moment and second formant frequency against the average MMN magnitude elicited by each of the repeated vowel deviants. Figure 3.8., demonstrates that vowels with a broader spectral bandwidth (or greater SM2) and higher F2 elicited MMNs of greater magnitude.

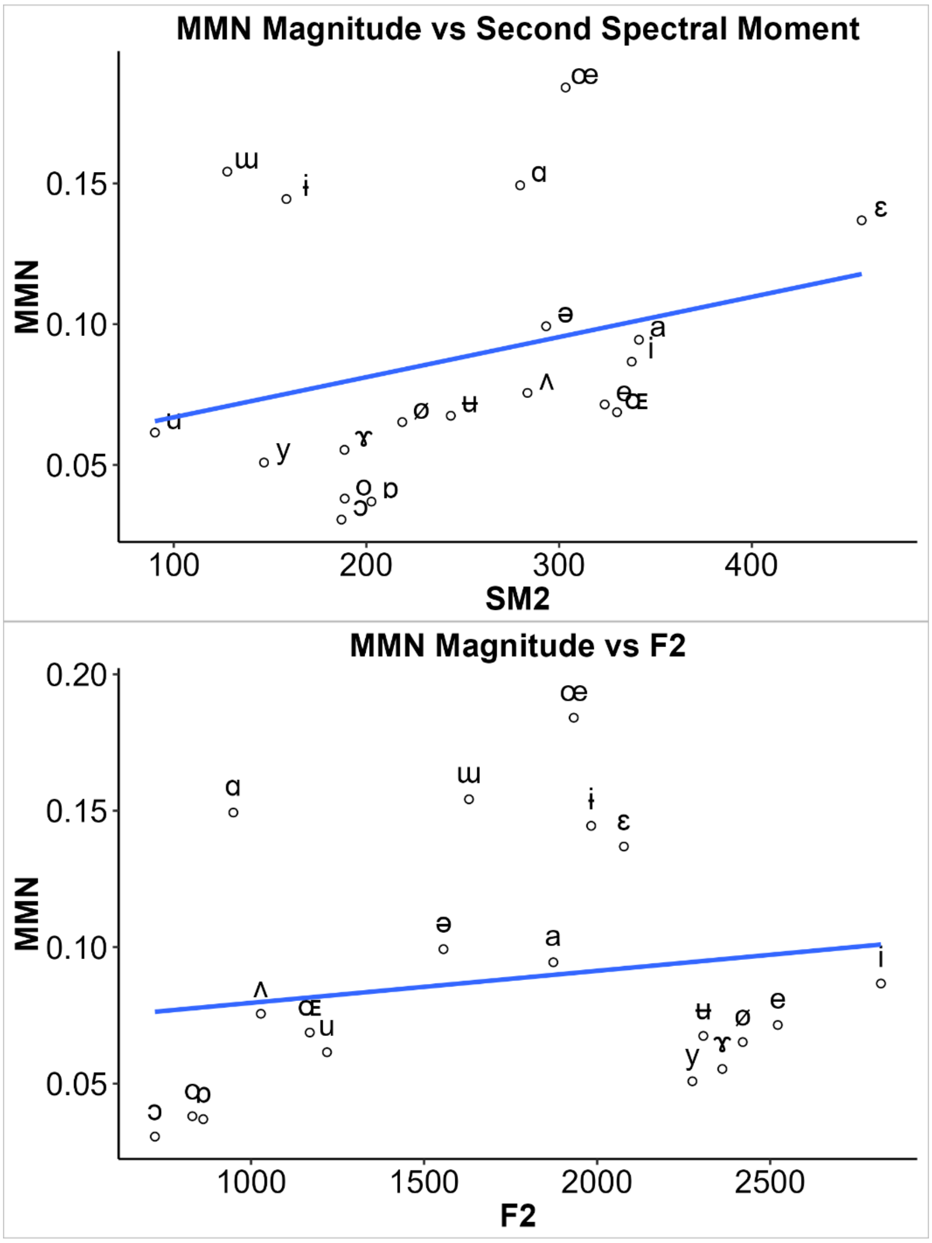

Figure 3.8. Scatterplots of SM2 and F2 against average magnitude of the MMN elicited by each IPA vowel deviant. Vowels with a broader spectral bandwidth and higher F2 elicited MMNs of greater magnitude.

3.4. Discussion

The current study tested a novel framework for MMN testing which allowed us to record perceptual indices to 19 IPA vowels in a single experiment. Previous MMN investigation had examined the perception of few phonetic contrasts because including more deviant types in the paradigm would lead to a decreased reliability of the recorded MMNs. That is, since the MMN is a difference wave and it is small in magnitude, ERP analyses require that many deviants are included in the paradigm, per deviant type tested, to conclude whether the difference wave is significant. However, the current experiment has demonstrated that using multivariate Temporal Response Functions (mTRF; Crosse et al., 2016), we are able to bypass this issue, as these functions gather more statistical power in the analysis. This increased power is achieved by using data from the 64 EEG channels whilst accounting for multicollinearity. Furthermore, permutation analyses of the raw EEG data and the mTRF data showed that for the 19 different deviant types, and with 7 standard responses per deviant response, the MMNs observed using the mTRF method were just as reliable after a single deviant as they were after 100 deviants, per deviant type, in a traditional ERP analysis. This is a remarkable finding since this methodological framework allows us to include more deviant types in a single oddball paradigm and test perception of multiple phonetic contrasts in a time-effective way.

The mTRF framework has been mostly used in neural tracking of the speech envelope (e.g. Attaheri et al., 2022; Zinszer et al., 2022) but some investigations have applied them to ERP components such as the N400 (e.g., Song & Iverson, 2018). A key difference between previous mTRF analyses and the current one is that we did not use the acoustic envelope as the stimulus input function. Instead, we created a function that would reflect differences between responses to standard and deviant stimuli, in a categorical manner. The standards were assigned even weights across time; whilst the deviants had function weights peaking upon onset of the deviant stimulus. The graphical representation of the averaged back-projected data, was similar to the graph displaying the input stimulus function i.e., the deviant-stimulus function had a sharp negative peak at time 0s and, from time 100ms, the deviant and standard functions showed similar function weights. On top of this, the ERP analyses also revealed increased negativity at around 200ms upon the onset of the deviant stimulus. Therefore, we show evidence of an MMN in the mTRF and the ERP analyses and demonstrate that mTRFs are appropriate for analysing MMN data.

In the current study, backwards models of the mTRF were preferred over forward models. The primary advantage of backwards models is that they do not necessitate the pre-selection of neural channels. Instead, each channel is automatically weighted based on how informative it is for stimulus reconstruction (Crosse et al., 2021). This approach is beneficial because we did not have any preconceived notions about which neural channels would be most relevant for reconstructing the stimulus. By using backwards models, we allowed the mTRF analysis to determine the most relevant channels. Additionally, backwards models have the capability to utilize independent information from various neural channels to extract features of the stimulus (Crosse et al., 2021). This means that the decoding functions can integrate diverse neural signals to infer the characteristics of the stimulus more effectively. Another significant advantage of backwards models is their typically higher reconstruction accuracy (Crosse et al., 2021). This is because decoders in backwards models project directly to the stimulus domain, providing more direct access to the actual source of changes in the stimulus, as opposed to dealing with the noisy data often encountered in EEG recordings. Lastly, decoders in backwards models can leverage any neural information that correlates with the stimulus feature, even if that neural activity does not explicitly encode the feature (Crosse et al., 2021). This makes the decoding process more robust and comprehensive, as it can incorporate a wider range of relevant neural signals.

To further refine mTRF analysis for MMN, future investigations should strive to investigate how much MMN paradigms can be shortened whilst still gathering enough EEG data to build reliable mTRF models. The developers of these machine learning functions, state that it is important to consider how much training data is required to avoid overfitting the model to specific features, or even noise, in the EEG data (Crosse et al., 2021). For most purposes, the general recommendation is recording a minimum of 10 to 20 min of data per condition (Crosse et al., 2021). Although collecting larger amounts of EGG data is preferred for increasing the model's accuracy, applying the current technique to infant testing would require shortening the experimental paradigm. As a result, the next step would be to explore the minimum amount of standard and deviant epochs needed to effectively train and test the mTRF models.

In the current paradigm, the brain likely builds expectations based on the regularities of the vowel pairs. Specifically, across standard trials, the auditory system learns to anticipate that the first vowel will differ from the second. This forms a predictive template in which variation in the second position becomes expected. When a deviant pair is presented (i.e., the same vowel is repeated), it violates this learned pattern, and an MMN response is elicited. Crucially, this MMN cannot be fully explained by neural adaptation, as no single vowel is repeated often enough to cause reduced neural responsiveness purely through habituation (see Näätänen et al., 2004). Instead, the MMN response to the deviant reflects a violation of the brain's prediction about the relationship between the two vowels. That is, that the two vowels should differ. This supports the predictive coding hypothesis (e.g., Wacongne et al., 2012), which posits that MMN arises from a mismatch between incoming auditory input and top-down predictions formed by the brain's internal model of the environment. By designing the standard sequence to contain variability and the deviant to be acoustically more redundant, this paradigm provides strong evidence that MMN reflects a prediction error, not merely differential adaptation to repeated sounds.

A secondary aim of our investigation was to record MMN responses that would index vowel categorization. This would allow us to build perceptual maps of the vowel space for each of our subjects, similarly to McCarthy et al., (2019). Our data suggests that our MMNs were reflective of phonetic processing, i.e., our ERP analyses revealed scalp topographies with more left-lateralised MMN sources than those reported in previous studies (Näätänen et al., 2007, 2017). This could be because our deviant stimuli did not entail an acoustic change and thus, MMN sources linked to acoustic change detection might not have been as strongly activated here. However, although the scalp topographies indicated effects of phonetic processing, analysis of the MMN magnitudes and identification responses demonstrated that the recorded MMNs were not accurate indices of vowel categorisation.

We hypothesised that the IPA-vowel deviants that participants perceived as native vowels would elicit MMNs of larger magnitudes but MMN magnitude was not related to vowel identification scores. The vowels with the greatest MMNs were /ɨ/, /œ/ and /ɯ/ whilst those that were consistently mapped onto a single native vowel category in the behavioural task were /i/, /u/, /ɑ / and /ʌ /. The vowels with higher identification scores were all native to southern British English listeners but surprisingly, the larger MMNs were recorded to non-native vowels. Potentially, the recorded MMN magnitudes reflecting lower-level auditory processing rather than phonetic processing. This is surprising, considering the left-lateralised source localisation patters. Nevertheless the mixed model analysis showed that the pattern of MMN magnitudes is not completely random.

Two acoustic properties of the vowels led to larger MMNs: higher second spectral moments (SM2) and higher second formant frequencies (F2). The literature indicates that "focal vowels" (i.e., those with lower second spectral moments) are perceived as more salient (Masapollo et al., 2017) and potentially, serve as better vowel exemplars. Consequently, the recorded pattern of MMN magnitudes suggests that the better vowel exemplars elicited smaller MMNs. Perhaps, non-native vowels with wider spectral distances (i.e., /ɨ/, /œ/ and /ɯ/) might not have been perceived as true vowels in the vowel-pair oddball paradigm. Deviants which included /ɨ/, /œ/, and /ɯ/ may have induced greater dissonance between the expected and the actual stimulus. In any case, the acoustic analysis was only carried out to identify regularities in the MMN magnitudes and our aim remains to create a paradigm that will elicit MMNs that reflect vowel categorisation.

One reason why the MMN magnitudes did not align with phonological categorization could be that the deviants in our paradigm did not entail a phonetic difference. The standard stimuli in our paradigm were vowel pairs composed of two different vowels. When a deviant was introduced, subjects heard two identical vowels, meaning that the deviant stimulus belonged to the same category as the vowel they had just heard. Studies that show greater MMNs for native vowels used deviants belonging to different phonological categories than the standard (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Sharma & Dorman, 1999). Therefore, the current vowel-pair oddball paradigm may not be well-suited for eliciting MMNs that reflect vowel categorization. An alternative would be to have same-vowel pairs as the standards and different vowel pairs as the deviants. This updated paradigm would continue to incorporate multiple vowels, which would serve as both standards and deviants at various time points. This design would allow for the measurement of perception across the 19 IPA vowels, ensuring that the deviant stimuli entail a phonetic distinction.

Additionally, our choice of stimuli may not have been appropriate for observing vowel categorization, as few of the IPA vowels were perceived as native. None of the IPA vowels were consistently assigned to a particular native category, and the category goodness ratings for all the vowels were quite low. Furthermore, some IPA vowels were not perceived as expected (i.e., the most common response for /e/ was 'beat' instead of 'bet'). Therefore, the IPA vowel stimuli were not good representations of SBE vowels. Ylinen et al., (2010) argued that the MMN's enhanced sensitivity to native vowel sounds is mostly observed when the native vowel exemplars are prototypical exemplars in the listener's native language. They argued that these prototypical exemplars create the strongest memory traces. The IPA vowels used in the current experiment were acoustically different to prototypical southern British English vowels, hence why our pattern of results does not show larger MMNs for those IPA vowels that are present in the southern British English repertoire. Using naturalistic recordings from native English speakers could determine whether the MMNs elicited in this paradigm reliably index vowel categorization, potentially leading to MMNs of greater magnitude for native vowels.

All in all, it can be concluded that the combination of a vowel-pair oddball paradigm and the mTRF analysis framework is a time-effective way of recording MMNs to 19 IPA vowel contrasts. In future experiments, the paradigm could be reversed to include same-vowel pairs as standards and different-vowel pairs as deviants, to elicit MMNs that will reflect phoneme categorization. Once the MMNs index categorisations, perceptual maps could be built for sets of phonemes. This would allow us to draw direct comparisons in their perception and said tool could allow us to explore how phonetic categories are built and develop throughout the lifespan.

4. Mismatch Negativity Framework for Cross-Language Vowel Perception.

4.1. Introduction

In adulthood, native phonetic categories influence how we learn unfamiliar sounds in a second language (L2). For example, learners often assimilate L2 phonemes that resemble native sounds into familiar categories (Best & Tyler, 2007). This similarity can support perceptual learning by helping learners map new sounds onto existing categories, which makes them easier to recognize and process. However, if the L2 contains two distinct categories that are perceived as equivalent to a single native category, learning to discriminate between these L2 contrasts becomes challenging (Best & Tyler, 2007). Therefore, studying cross-language phonetic perception allows us to understand how native phonetic categories shape the perception of unfamiliar L2 sounds. Cross-language phonetic perception has traditionally been studied using behavioral tasks that prompt listeners to identify different phonemes (Best, 1995; Flege, 1995; Iverson & Kuhl, 1995). Although these tasks reveal how listeners assimilate and categorize phonemes, they do not clarify whether responses stem from automatic processing or conscious judgments. In contrast, neural measures offer a way to observe how the brain automatically processes speech sounds, without the influence of conscious decision-making.

The current study uses the MMN response as a tool to investigate cross-language vowel perception differences, using a vowel-pair oddball paradigm with 19 IPA vowel contrasts. As mentioned in Chapter 3, native phonemic contrasts elicit larger MMNs than similar non-native or non-phonemic differences (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Sharma & Dorman, 1998, 2000). Later studies showed that even native sound pairs elicit different neural responses depending on their phonemic status. For example, Kazanina et al., (2006) compared MMN responses to deviant [t] and [d] sounds in Russian and Korean speakers. In Russian, these two sounds belong to different phonological categories but in Korean, they are allophones of the same phonological category. Russian speakers showed a clear MMN response to [d] and [t] sounds, indicating automatic phonemic categorization, whereas Korean speakers did not exhibit a significant MMN, reflecting the lack of phonemic distinction for these sounds in Korean. Moreover, Hisagi et al., (2010) showed that the unique phonological features of the sound can influence the MMNs that it will generate. They recorded MMNs to vowel length contrasts in naïve American English speakers, for whom vowel duration is not a primary phonemic cue and Japanese speakers whose native repertoire includes long and short vowels. Their results demonstrated significantly larger MMNs to vowel-length deviants in the Japanese speakers than in Americans, demonstrating MMN sensitivity to native phonological features. Chládková et al., (2013) extended these findings by comparing Spanish and Czech listeners. Spanish speakers, whose language does not use vowel length contrastively, showed larger MMNs to duration changes in non-native vowels than in native ones. In contrast, Czech speakers, whose language encodes vowel duration phonemically, showed large MMNs for native vowel-length changes but smaller responses to non-native contrasts.

The current study uses the MMN response to explore vowel perception in Southern British English speakers and Spanish-native learners of English. In addition to its role in cross-language phonetic perception, researchers have used the MMN to study how the brain learns new speech sounds in L2 learning. According to current theory, the MMN is sensitive to native contrasts because speech perception is more or less categorical: memory traces for native sounds and especially prototypical native sounds amplify the MMN response beyond what would be expected from acoustic differences alone (Shtyrov & Pulvermüller, 2007). As L2 learners gain experience, they may begin to form memory traces for L2 phonemes as well.

With sufficient exposure and brain plasticity, their neural responses to L2 sounds could start to resemble native-like patterns, potentially leading to stronger MMN responses.

Nonetheless, MMN investigations in L2 learners do not show straightforward results. On the one hand, Winkler et al., (1999) demonstrated that Hungarian adults who were late learners of Finnish and had been immersed in their L2 environment for several years, showed similar MMN amplitudes to Finnish speakers to a Finnish vowel contrast. However, this finding was not replicated by Peltola et al., (2003) who found that advanced Finnish students of English did not show MMN responses to English phonemes comparable to those elicited by native Finnish phonemes. Nonetheless, these advanced Finnish students of English did not have equivalent levels of immersion in their L2 to the participants in Winkler et al., (1999). Meaning that individual differences, such as L2 immersion experience, influence the development of MMN traces for L2 speech sounds. Additionally, researchers have explored whether length of in-classroom language instruction has an effect on the MMNs to target-language phoneme contrasts. Grimaldi et al., (2014) tested two groups of University students in their first and fifth years of English instruction and found that the MMNs to English L2 vowel deviants were not significantly different to those elicited in naïve listeners of English. Additionally, Wottawa et al., (2022) tested French learners of German on native German vowel and consonant contrasts and found that although German native speakers demonstrated reliable MMNs to the vowel-length deviants, the French learners demonstrated much smaller negativities to the deviant stimuli. In a 19-month study, Højlund et al., (2022) found no significant increase in MMN responses to Dari and Arabic phonetic contrasts among two groups of learners, as their learning progressed. Therefore, it appears that in-classroom language learning may not be enough to influence the pre-attentive L2 phoneme discrimination process that are reflected in MMN responses.

However, MMN responses to non-native phonemes can be modulated through short-term phonetic training, demonstrating that the MMN can reveal effects of phonetic learning. For instance, Tremblay et al., (1998) investigated native English speakers' MMN responses to a voicing onset time (VOT) contrast (/ba/ vs. /mba/). While participants initially perceived the stimuli as identical, post-training MMNs indicated neural discrimination of the sounds, despite their inability to overtly identify the feature distinguishing the stimuli. This suggests that speech sound learning can occur at a pre-attentive level. Similarly, Zhang et al., (2009) demonstrated enhanced neural sensitivity to the English /r-l/ contrast in Japanese adults following training, while Ylinen et al., (2010) found that Finnish speakers trained to discriminate the English /i/-/ɪ/ contrast showed improved pre-attentive processing of spectral cues, as reflected in post-training MMN responses. A caveat of these studies is that the potential long-term effects of these training protocols remain largely unexplored, with the exception of Kraus et al., (1995), who demonstrated sustained neural and behavioral changes in perception of similar speech tokens one month after the training period. Nevertheless, the reviewed evidence suggests that MMN elicitation is modulated by the degree of familiarity with the phonemic repertoire of a given language, making MMNs an appropriate tool for studying phonetic perception and its development.

The current study aims to use MMN amplitudes to build perceptual maps of the vowel spaces in southern British English speakers and Spanish speakers. Spanish has five vowel phonemes, which contrast along two dimensions: vowel height and front/backness of the tongue position. There are two high vowels /i/ and /u/, two mid-vowels /e/ and /o/, and one low vowel /a/. Along the front/back dimension, /i/ and /e/ are the two front vowels, /a/ is a central vowel and /u/ and /o/ are back vowels. These vowels do not alternate as a function of lexical stress (Simonet & Amengual, 2020). Standard Southern British English, a recognized variety of Southern British English, has a vowel system comprising 11 monophthongs: /iː, ɪ, ɛ, ɜː, a, ɑː, ʌ, ɒ, ɔː, ʊ, uː/, excluding /ə/, which appears only in unstressed syllables (de Jong et al., 2007). In addition to Spanish, English vowels can also be categorised in terms of lip rounding and, as either tense or lax vowels. Tense vowels, such as /iː/ and /uː/, are typically longer in duration and involve more muscular tension, whereas lax vowels, such as /ɪ/ and /ʊ/, are shorter and involve less tension.

Along the height dimension, SSBE can be divided into four high vowels, /iː, ɪ, uː, ʊ/; five mid vowels /ɛ, ɜː, ə, ʌ, ɔː/; and three low vowels /a, ɒ, ɑː/. Along the front-back dimension, there are four front vowels /iː, ɪ, ɛ, a/; five back vowels /uː, ʊ, ɔː, ɒ, ɑː/; and three central vowels /ɜː, ə, ʌ/. Additionally, rounded vowels include /uː, ʊ, ɔː, ɒ/, while unrounded vowels include /iː, ɪ, ɛ, ɜː, a, ə, ʌ, ɑː/ (Collins & Mees, 2013). All together, the vowel inventories of the two speaker groups reveal large differences, with only the phonemes /i/ and /u/ demonstrating cross-linguistic similarity. As a result, the perceptual maps are anticipated to differ substantially, as they are inherently shaped by the native vowel inventories of the speakers.

To elicit MMNs, we use a similar oddball paradigm to that described in Chapter 3. The same 19 IPA vowels are arranged into standard and deviant vowel pairs with standard pairs consisting of a repeated vowel pair (i.e., a-a, a-a, a-a) and two deviant types going in parallel. The series deviant breaks the sequence of repeated vowels with a repetition of a different vowel (i.e., a-a, a-a, e-e) and the pairwise deviant contains a different second vowel (i.e., a-a, a-a, a-e). In the oddball paradigm, 15% of tones are deviant vowel pairs, picked randomly, and each deviant is always followed by at least one standard pair. The decision to modify the standard and deviant pairs was driven by the observation that previous studies reporting larger MMNs for native speech sounds often used deviant stimuli that differed acoustically and phonetic/phonologically for at least one of the speaker groups (Chládková et al., 2013; Hisagi et al., 2010; Kazanina et al., 2006; Näätänen et al., 1997). If MMN magnitudes are consistently larger for native vowels, they may reveal cross-language effects between English and Spanish speakers, reflecting their distinct processing of the vowel pairs.

The Spanish-speaking group in the current investigation lived in London at the time of test and their knowledge of the English vowel repertoire was assessed with an auditory word recognition task in which they listened to various words and selected the perceived word from five possible options. We hypothesized that the English phonetic knowledge of the Spanish-speaking group would not cause their MMN responses to resemble those of English speakers, as previous research has not consistently demonstrated MMN adaptation to such neural adaptations to non-native speech sounds (Grimaldi et al., 2014; Højlund et al., 2022; Peltola et al., 2003; Wottawa et al., 2022).

To investigate cross-language effects, a trial-by-trial mixed model was constructed, with language group coded as a fixed effect. As part of the data preprocessing, the EEG data was de-noised using backward multivariate Temporal Response Functions (mTRFs). This method was selected based on the findings from Chapter 3, which demonstrated that mTRF-based back-projections generated neural responses closely aligned with the input function, confirming the suitability of this technique for analyzing MMN responses. Lastly, to examine the relationship between MMN amplitudes and phonetic assimilation, identification (ID) data from English speakers (reported in Chapter 3) were analyzed, and additional ID scores were collected for Spanish speakers using a similar identification task. These assimilation scores were then related to the MMN magnitudes in a second mixed model, allowing for an investigation into cross-language variability in neural responses to the 19 IPA vowels.

4.2. Methods

4.2.1. Subjects

A total of 34 participants took part in the experiment. Out of these, 17 were native British-English adults (M=28.86 years, SD= 11.14 years). Data from one of these subjects was excluded from the analyses because it was too noisy. The remaining 17 subjects were Spanish-speaking adults (M=33.14 years, SD= 7.02 years) who lived in London and had learned English as a second language in late childhood or adulthood. Participants were students or staff members at University College London (UCL) and were contacted through the UCL Psychology Subject Pool.

4.2.2. Materials

We designed three tasks for this experiment. First, we created a vowel-pair oddball paradigm for both Spanish and English speakers. In addition, we designed two behavioural tasks of vowel perception for the Spanish-native participants: a vowel identification task and category goodness task and a perceptual test of southern British English vowels. The vowel identification task was the same task as the one used in Chapter 3, adapted for Spanish speakers. The purpose of the English vowel test was to confirm that the Spanish speakers had not acquired native-like perception of English vowels.

4.2.2.1. Oddball Paradigm

To elicit MMNs, we employed an oddball paradigm similar to the one described in Chapter 3. The same set of 19 IPA vowels was used to construct standard and deviant vowel pairs. Standard trials consisted of repeated vowel pairs (e.g., a-a, a-a, a-a). Two types of deviants were presented in parallel: series deviants, which broke the repetition pattern by introducing a different repeated vowel (e.g., a-a, a-a, e-e), and pairwise deviants, in which the second vowel differed from the first within the pair (e.g., a-a, a-a, a-e). Deviant trials made up 15% of all stimuli and were randomly distributed, with the constraint that each deviant trial was always followed by at least one standard pair. After a series deviant was introduced, the deviant repeated vowel pair became the new standard (e.g., a-a, a-a, **e-e**, e-e, e-e). One vowel pair was played every second with the second vowel in the pair starting 200ms after the first one, leaving a 600ms gap between pairs.

4.2.2.2. Vowel Identification and Category Goodness Task for Spanish speakers

We used a similar vowel identification (ID) and category goodness task as the one described in Chapter 3. The task employed the same 19 IPA vowels and was implemented using the Gorilla Experiment Builder. However, unlike the previous ID task, the response options in each trial remained constant, with participants choosing from the words *masa*, *mesa*, *misa*, *mosca*, and *musa*. These five words were selected because the vowel in the first syllable of each word corresponds to a vowel in the Spanish vowel repertoire. Participants heard one IPA vowel at a time and had to identify the word that contained the vowel sound, in its first syllable, that most closely resembled the IPA sound that they heard. Category goodness ratings were collected in a manner consistent with the procedure described in Chapter 3 i.e., by positioning a pointer along a spectrum of “very good” to “very poor” vowel exemplar.

4.2.2.3. Test of English Vowels

To assess the perceptual abilities of the Spanish speakers with regards to the southern British English vowel set, a short vowel test was included in the vowel identification and category goodness task. In each trial of this test, participants listened to an English minimal pair bVt word (i.e., bit, beat, bait…), and had to identify which word they heard, from 6 response options. There was a total of 14 minimal pair words included in the test and these are their IPA symbols: / i, ɪ, aɪ, eɪ, ɛ, ɑ, a, ʌ , ɒ, əʊ, ɔ, u, aʊ, ɜ /. Participants were prompted to identify each minimal pair 4 times.

4.2.3. Procedure

The Spanish participants completed the behavioural tasks online before taking part in the EEG study. To do this, they accessed the Gorilla Experiment platform and followed detailed instructions. They were asked to complete the task in a quiet space and to use headphones, whenever possible. In each round of the behavioural task, of which there were four, participants completed 19 trials of vowel identification and category goodness ratings, followed by 14 trials of the test of English vowels. During the vowel identification and category goodness task, they were prompted to "think in Spanish" and before the test of English vowels, they were warned that the task would switch to English.

Once they completed the behavioural tasks, they were invited to take part in an EEG recording in the laboratory. The English native participants only took part in the EEG-recording part of the experiment as we already obtained behavioural vowel judgements from English speakers in Chapter 3. Testing sessions were carried out in a sound-proof EEG booth where participants took part in a passive listening task whilst they watched muted cartoons. In the listening task, participants heard 12 sequences of the oddball paradigm, lasting two and a half minutes each. They heard a total of 1920 vowel pairs binaurally through Etymotic ER-1 insert earphones and they were allowed to take breaks in between sequences. The total EEG recording time was of around 45 minutes and each participant heard a unique set of oddball sequences.

4.2.4. EEG recording and analysis

The same EEG equipment and recording procedures were used, as those described in Chapter 3. EEG was recorded through a Biosemi Active Two system with 64 (Ag/AgCl) electrodes mounted on an elastic cap and 6 external electrodes: left and right mastoids and two vertical and horizontal EOG electrodes. Recordings were made with a sampling rate of 2048 Hz and downsampled to 64Hz.

EEG data were preprocessed using a custom MATLAB pipeline incorporating EEGLAB (Delorme & Makeig, 2004), FieldTrip (Oostenveld et al., 2011), and NoiseTools (de Cheveigné & Simon, 2008) functions. Recording blocks were identified and cut points were tapered using a 1-second Hanning window to minimize boundary artifacts. Vertical and horizontal EOG signals were computed by bandpass filtering the appropriate external channels. High-pass filtering at 0.1 Hz was performed separately within each recording block, using a zero-phase first-order Butterworth filter. Artifact rejection was conducted using Denoising Source Separation (DSS), targeting components most correlated with VEOG and HEOG signals; highly correlated components ($r > 0.2$, rank < 6) were removed to suppress blink and saccade-related artifacts. Finally, a zero-phase first-order Butterworth low-pass filter at 8 Hz was applied, and the data were downsampled to 32 Hz for further analysis.

4.3. Results

In the first instance, we analysed the Spanish speakers' scores on the test of English vowels and found that they were overall low, where the mean was 50% correct and the standard deviation was 0.09. This demonstrates that the Spanish and English participants had different vowels systems, although the Spanish speakers were living in an English-speaking country. Moreover, these results suggest that the MMNs elicited in English and Spanish speakers should reveal cross-language effects.

The EEG data was first analysed using an event-related potential (ERP) analysis to evaluate whether our updated version of the oddball paradigm elicited reliable MMNs in our participants. In this initial step, we were only interested in looking at the averaged data across our participant groups to observe potential differential responses to standard and deviant vowel pairs. In the second step, we applied multivariate temporal response functions (mTRFs) to de-noise the EEG, building on the findings from Chapter 3 that validated mTRFs for MMN analysis. This allowed us to examine cross-language effects on MMN magnitudes at the single-trial level. That is, we built two mixed-effects models using single-trial MMN data to test for effects of language background and phoneme identification. For this latter analysis, we used identification scores from Spanish-native participants in the current study and English-native scores from Chapter 3.

4.3.1. ERP analysis

In the first step, we used data from the Fz channel to carry out an average ERP for neural responses to standard, series deviant, and pairwise deviant vowel pairs, across all our subjects. Figure 4.1. shows the averaged ERP response for the standard responses, depicted in red, across all subjects. The graph includes a black line representing the amplitude envelope of the vowel pair used as the auditory stimulus, serving as a temporal reference.

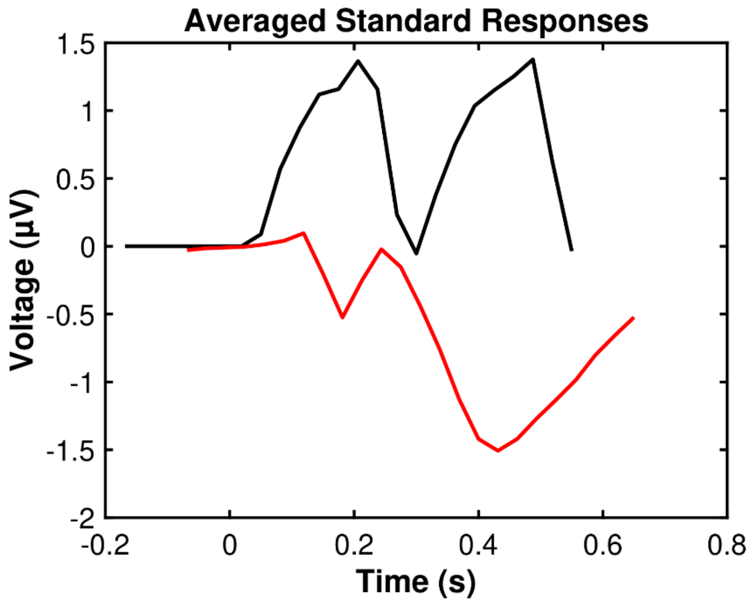

Figure 4.1. Averaged ERP Response for Standard Vowel Pairs. The graph displays the averaged ERP response in red, with a black line indicating the amplitude envelope of the vowel pair used as the auditory stimulus, serving as a temporal reference.

The averaged ERPs for standard pairs showed a classic P1, N1, P2 response for the first vowel in the pair. The second vowel did not show that classic response, probably because the stimulus is more complex after the onset of the second vowel. In Figure 4.2. the averaged standard ERP is depicted in red and the averaged series deviant ERP is depicted in blue. Figure 4.3. shows the averaged pairwise deviant ERP (in blue) and the averaged standard ERP in red. In both figures, the vowel-pair amplitude envelope is shown in black as a temporal reference.

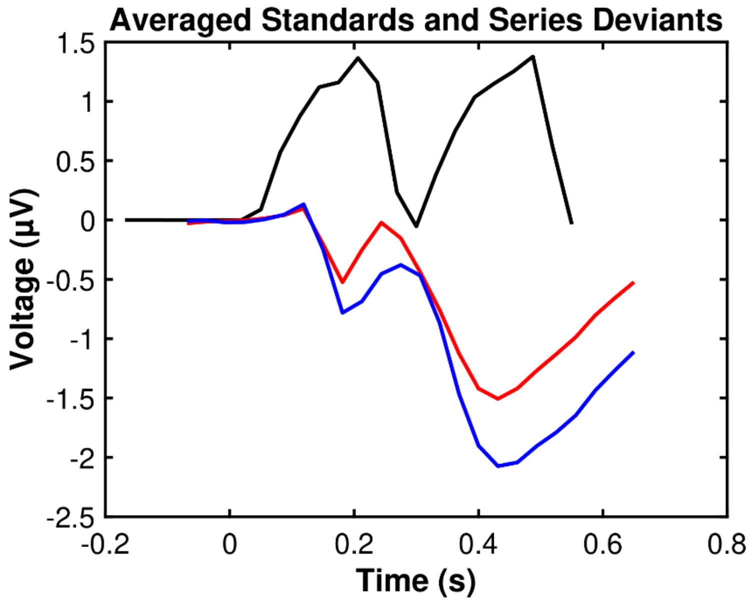

Figure 4.2. Averaged Standard and Series Deviant ERP Responses. The graph shows the averaged standard ERP response in red and the averaged series deviant ERP response in blue. The vowel-pair amplitude envelope is depicted in black, serving as a temporal reference.

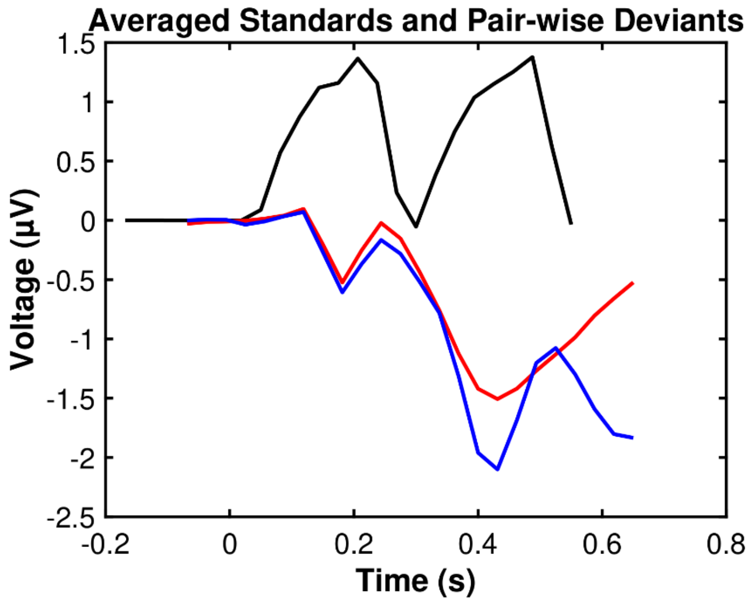

Figure 4.3. Averaged Standard and Pairwise Deviant ERP Responses. The graph displays the averaged standard ERP response in red and the averaged pairwise deviant ERP response in blue. The vowel-pair amplitude envelope is shown in black, serving as a temporal reference.

The averaged ERPs for the series deviants showed increase negativity, in relation to the standard response, after the onset of the second vowel, which persisted through time. In the pairwise deviant ERPs, where the deviation in the stimulus was specifically focused to the second vowel, the increased negativity upon onset of the second vowel was localised at around 200ms after the onset of the second vowel. Hence, visual inspection of these averages shows mismatch responses which look different for both types of deviant pairs.

Two sensor spaces were created: one for the series deviants and another, for the pairwise deviant responses. In each sensor space, the standard responses were subtracted from the respective deviant responses, allowing for visualization of the localization of electrodes where the most significant changes in voltage were recorded. Figure 4.4. and Figure 4.5. illustrate the sensors spaces for the series and the pairwise deviants, respectively, both displaying very similar results and revealing a fronto-central source of the MMN.

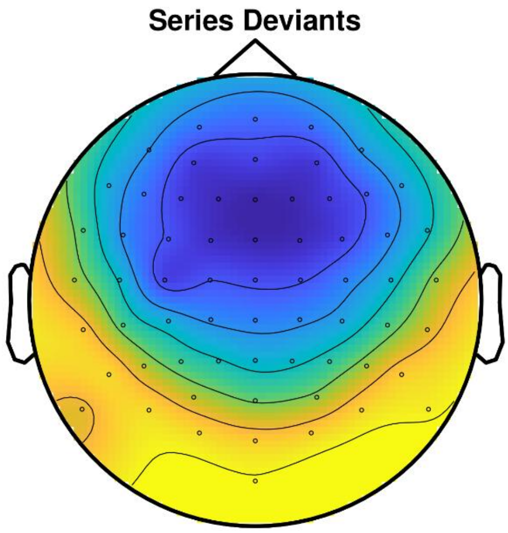

Figure 4.4. Topoplot of Series Deviants. The plot reveals a fronto-central source of the MMN, showing the areas of the brain where the most significant voltage changes were recorded for series deviants.

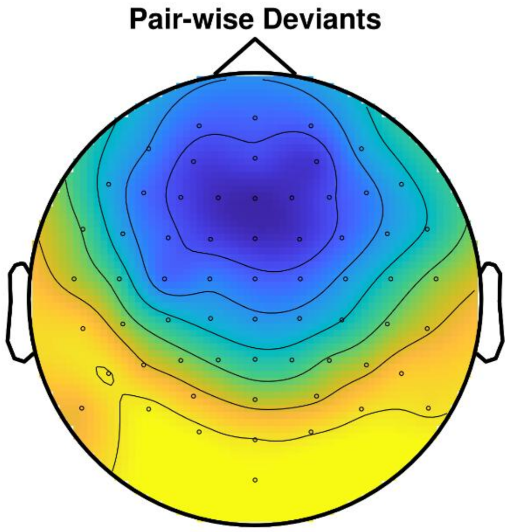

Figure 4.5. Topoplot of Pairwise Deviants. Similar to the series deviants, this plot also reveals a fronto-central source of the MMN, highlighting the consistency in the localization of the mismatch response.

4.3.2. mTRF analysis

The ERP plots demonstrated that our paradigm elicited an MMN response to the deviant vowel pairs. As a result, we aimed to determine whether the magnitude of the MMNs would be influenced by the language background of the participants (i.e., English vs. Spanish) and the identification scores of the IPA vowels. To explore these cross-language effects on the MMN we aimed to use mixed-effects model analyses with single-trial MMN data. Before building the models, we fit mTRFs to back-project the EEG data to the stimulus function, representing the speech sounds that participants heard during the experiment. This step allowed us to de-noise our EEG data.

The first mTRF analysis that we applied to the MMN data was the same as that carried out in Chapter 3. A stimulus function was created to assign different weights to the standard and the deviant responses. This stimulus function was used to generate a reconstructed stimulus function for each of our subjects, using regularised regression analysis and cross-validation. Similarly to Chapter 3, this analysis allowed us to effectively reconstruct the input stimulus function and confirmed that mTRFs are suitable for analysing the MMN data in this study.

In the following mTRF analysis stage, rather than building a stimulus function that would classify neural responses into standards and deviants, we calculated the amplitude envelope of an example vowel pair and used this amplitude envelope as the stimulus function in the model. Figure 4.6. shows the vowel-pair stimulus function created for the mTRF analysis. This function depicts the continuous, acoustic characteristics of the vowel-pair stimuli, enabling the mTRFs to model the brain's responses over time. It includes two amplitude peaks, with each peak representing a vowel sound in the pair.

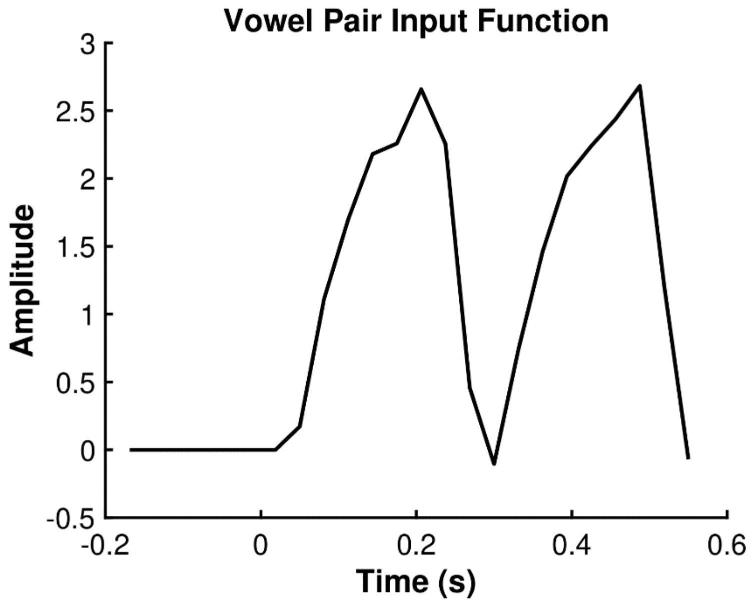

Figure 4.6. Vowel-pair Stimulus Function. The function, created for the mTRF analysis, includes two amplitude peaks, each representing a vowel sound in the pair.

For each subject, we back-projected the neural data from all 64 channels onto the input function in Figure 4.6 to obtain a reconstructed stimulus function, using mTRFs (Crosse et al., 2016). To evaluate the suitability of mTRFs, we plotted out the averaged back-projected standard, series deviant and pairwise deviant data. Visual inspection of the back-projected standard responses showed that this backwards projection framework effectively reconstructed the amplitude peaks in the envelope. Figure 4.7. shows the averaged back-projected standard responses in red and for reference, the vowel-pair stimulus function is illustrated in black.

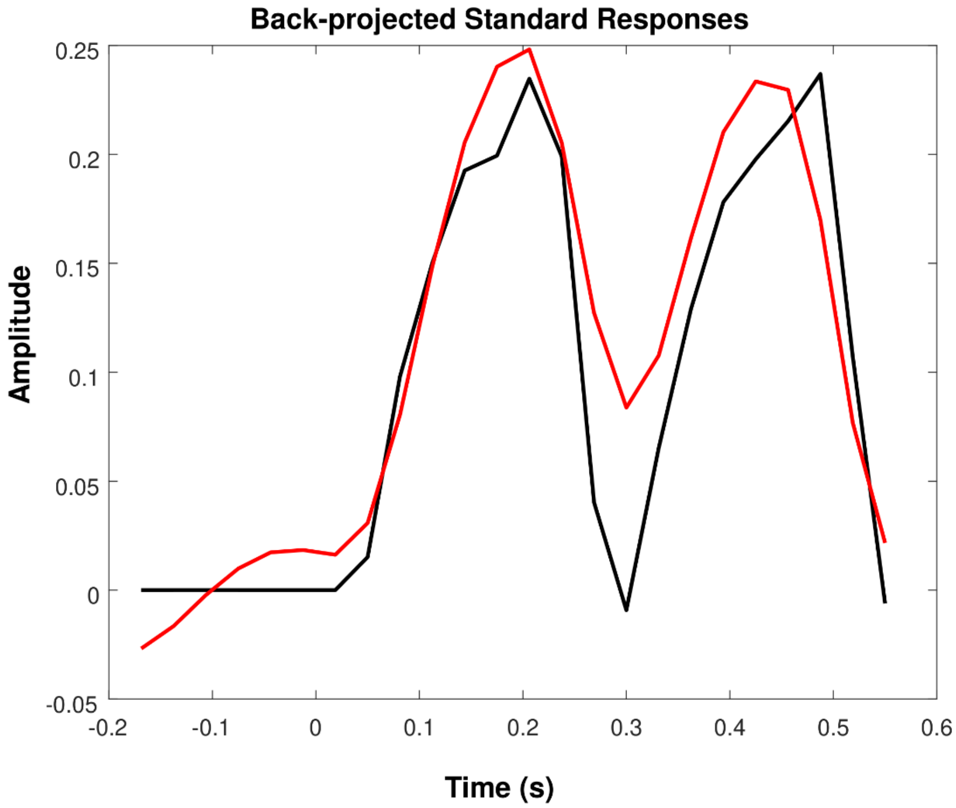

Figure 4.7. Averaged Back-projected Standard Responses. Averaged back-projected standard responses (red) compared to the vowel-pair stimulus function (black). The plot illustrates the temporal alignment between the neural responses and the stimulus. The back-projected standard responses shows two positive amplitude peaks, similarly to the vowel-pair stimulus function.

Moreover, the back-projected deviant data revealed a different pattern to the standards. Figure 4.8. and Figure 4.9. show the back-projected series and pairwise deviants, respectively as compared to the back-projected standards and for reference, the vowel-pair stimulus function is illustrated in black.

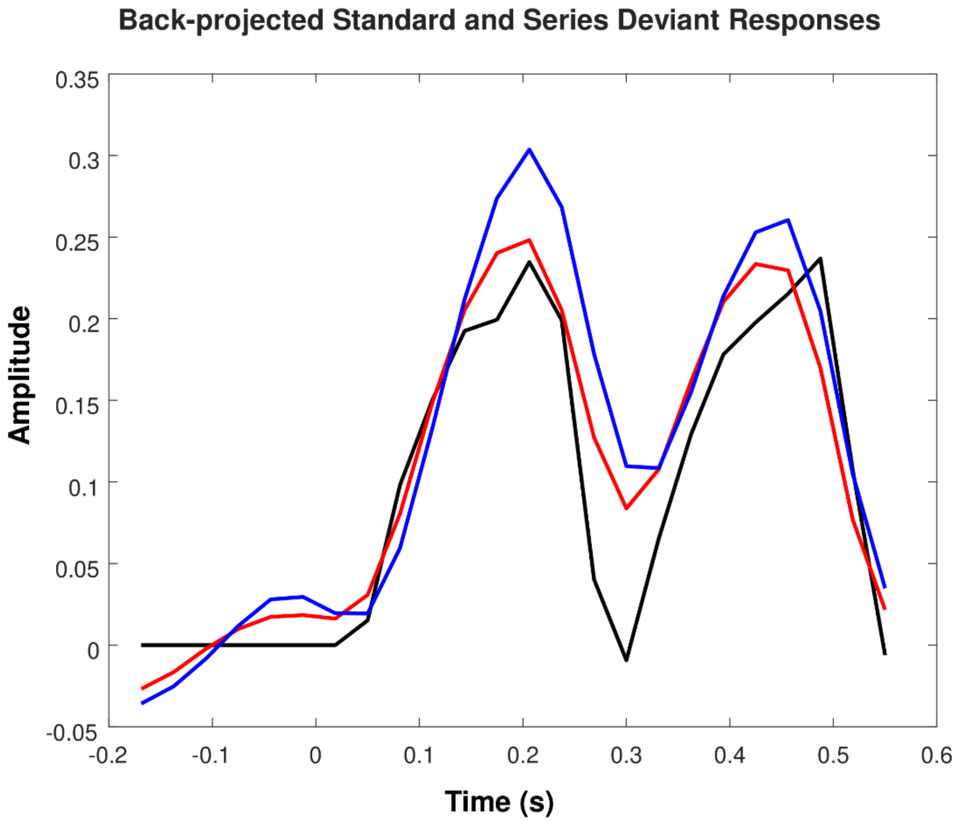

Figure 4.8. Averaged back-projected series deviants (blue) compared to the standard responses (red). The plot shows increased positivity in the first vowel, with a continuing but lesser positivity for the second vowel in the pair.

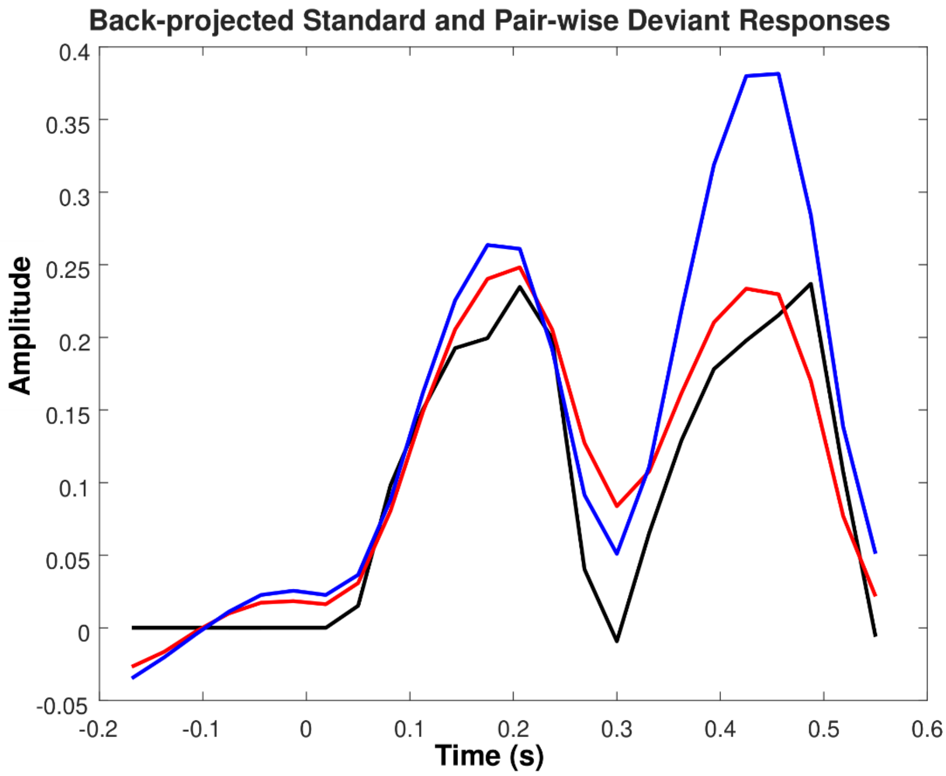

Figure 4.9. Averaged back-projected pairwise deviants (blue) compared to the standard responses (red). The plot illustrates a first amplitude peak similar to the standards and a significantly larger negativity 200 ms upon onset of the second vowel, displaying a precise localization of the mismatch response in the second vowel.

In both deviant types, the back-projected mismatch reponse was translated into increased positivity rather than negativity. The averaged back-projected series deviants showed increased positivity in the first vowel, as compared to the standards, with a continuing though lesser positivity for the second vowel in the pair. The averaged back-projected pairwise deviants displayed a first amplitude peak similar to that of the standards and a much bigger increase in positivity in the second vowel, as compared to the standards and the series deviants. Hence, the localisation of the mismatch response to the second vowel of the pairwise deviants is also present in the back-projected data.

4.3.3. Language Effects in MMN magnitudes

Since Figure 4.8 and Figure 4.9 demonstrate that the back-projected data highlights an MMN response to deviant pairs, we used the back-projected data to identify the factors influencing the MMN magnitudes. Firstly, we built a model to observe effects of language background on the MMN. The statistical analyses reported below were conducted using mixed models with the lme4 package (Bates et al., 2015)in R. A stepwise procedure was employed to remove non-significant factors (anova() comparisons, $p > 0.05$), and the significance of factors was determined using the CAR package (Fox & Weisberg, 2011) to obtain p-values within a type II analysis of deviance table and also, through model comparisons. Additionally, significance tests for individual factors and interactions were performed using the emmeans package (Lenth et al., 2025). All fixed and random effects are detailed below.

The first objective was to determine whether the native language of the listener influenced the amplitude of the back-projected MMNs. The mixed-effects model included a fixed effect of listener's native language, which was a categorical factor with two levels, English vs. Spanish. The current and previous stimulus were also included as categorical fixed effects, each containing 19 levels, corresponding to the 19 IPA vowels. The acoustic distance between the current and previous stimuli was included as a continuous fixed effect. Lastly, the type of deviant heard was also included as a fixed effect with two levels, series vs. pairwise deviants. Additionally, a factor of subject was included as a random effect to account for individual variability. The model's dependent variable was the peak amplitude of the stimulus reconstruction, measured for each deviant stimulus and each subject. For the series deviants, the amplitude peak was extracted from the 400 ms time window. For the pairwise deviant, it was extracted from the final 200 ms of the stimulus, as the deviance did not occur in the initial part. This means that single-trial deviant data was analysed for every subject and only the deviant responses were included in the statistical analyses.

To determine the most appropriate predictors in the model, we conducted an Analysis of Deviance using Type II Wald chi-square tests. The only significant main effect was that of deviant type (χ^2 (1) = 3.98, p = 0.04). The effects of the current stimulus on the back-projected MMN amplitudes interacted with the effects of the subject's language background (χ^2 (18) = 35.85, p = 0.007). The effects of the previous stimulus also interacted with the effects of language background (χ^2 (18) = 30.05, p = 0.03), as did the effects of the deviant type (χ^2 (1) = 10.05, p = 0.002). However, the acoustic distance predictor did not show any significant main or interaction effects.

To identify the best-fitting model for predicting the back-projected MMNs, we compared this full model, with a reduced model that excluded the acoustic distance predictor. The full model had 80 parameters (k=80), with an Akaike Information Criterion (AIC) of -3752, Bayesian Information Criterion (BIC) of -3106, log-likelihood (log$L$) of 1956, and deviance (D) of -3912, while the reduced model had 78 parameters (AIC = -3756, BIC = -3126, log$L$ = 1956, D= -3912). The reduced model had lower AIC and BIC values, indicating a better fit. An ANOVA comparing the two models showed that the addition of the distance predictor did not significantly improve the model fit (χ^2(2) = 0.66, p = 0.72). Therefore, the reduced model, which excluded the acoustic distance predictor, was selected as the final model.

In the final model, Type II Wald Chi-Square tests revealed a significant main effect of deviant type (χ^2(1) = 3.98, p = 0.046) whereby the pairwise deviants yielded greater MMN amplitudes (M=0.024, SD=0.29) than the series deviants (M=0.018, SD=0.15). The effect of deviant type on the amplitude of the back-projected MMNs showed a significant interaction with the effects of language background of the subjects (χ^2(1) = 10.04 p = 0.002).

To further explore this interaction, the *emmeans* package was used to obtain estimated marginal means and perform pairwise comparisons. These revealed that the interaction between deviant type and language background was driven by a significant difference in the Spanish group, whereby pair deviants (EMM=0.033) elicited greater MMNs than series deviants (EMM=0.019) with a difference of 0.014 (SE= 0.004, z= -3.64, p= 0.0016).

Moreover, in the final model, the effects of the current stimulus also showed a significant interaction with the language background of the subject ($\chi^2(18) = 35.37$, $p = 0.008$) and there was a significant interaction between the previous stimulus heard and language background ($\chi^2(18) = 29.51$, $p = 0.042$). The *emmeans* package was used again to unpick these interactions but in this case, despite multiple pairwise comparisons, none of the interactions reached statistical significance. As a result, we explored these interactions graphically to better visualize potential patterns and relationships between the language background of the subjects and the back-projected MMN amplitudes elicited by each IPA vowel. Figure 4.10. and Figure 4.11. show the average amplitudes of the back-projected MMNs elicited by each of the 19 IPA vowels in the series deviants for the English and the Spanish participants, respectively.

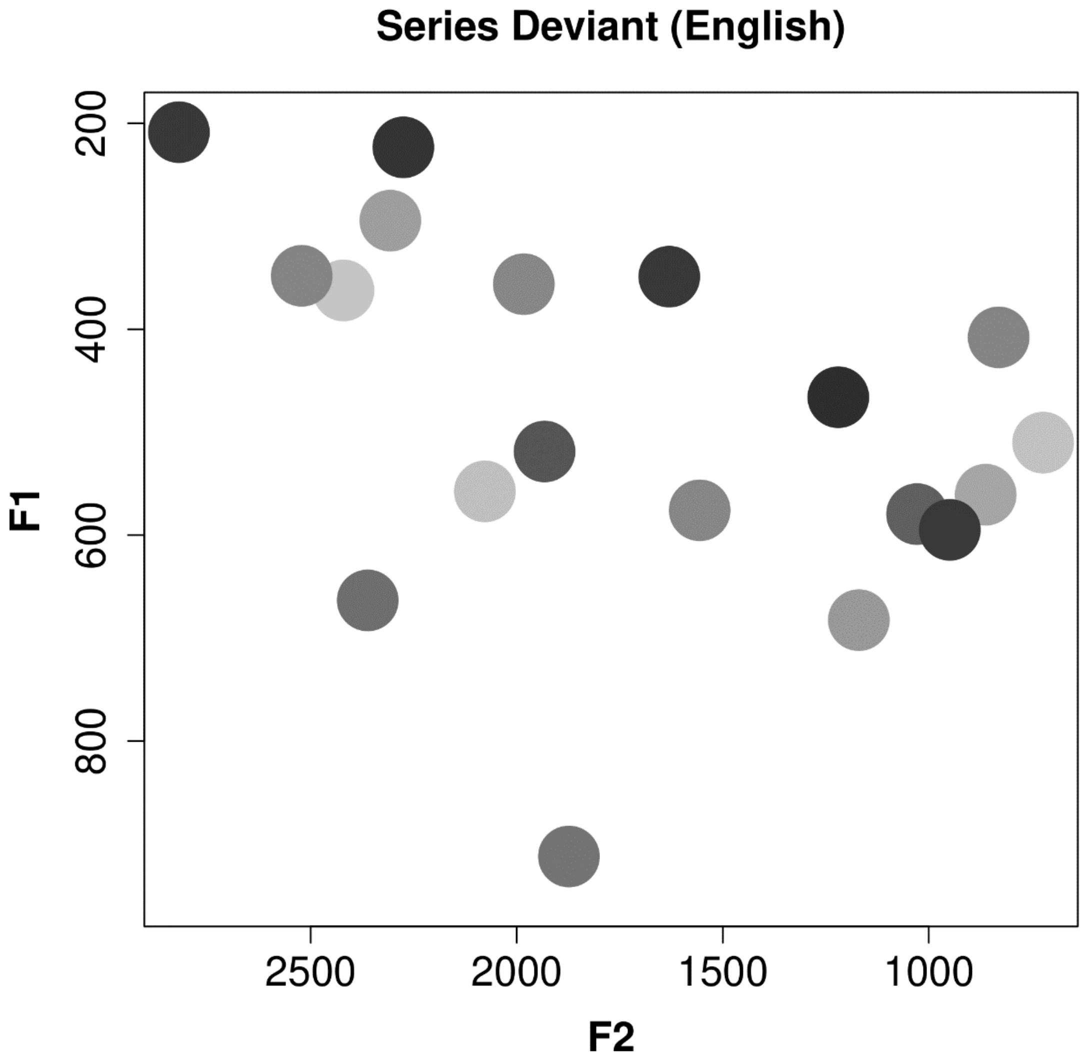

Figure 4.10. Vowel Plot of MMN Amplitudes for Series Deviants in English Participants. Each dot represents an IPA vowel and the darkness of each dot indicates the MMN amplitude that it generated. The MMN amplitudes were calculated by subtracting the average peak amplitude of the back-projected standard data from the average peak amplitude of the back-projected deviant data for each vowel. The darker the dot, the greater the average back-projected amplitude.

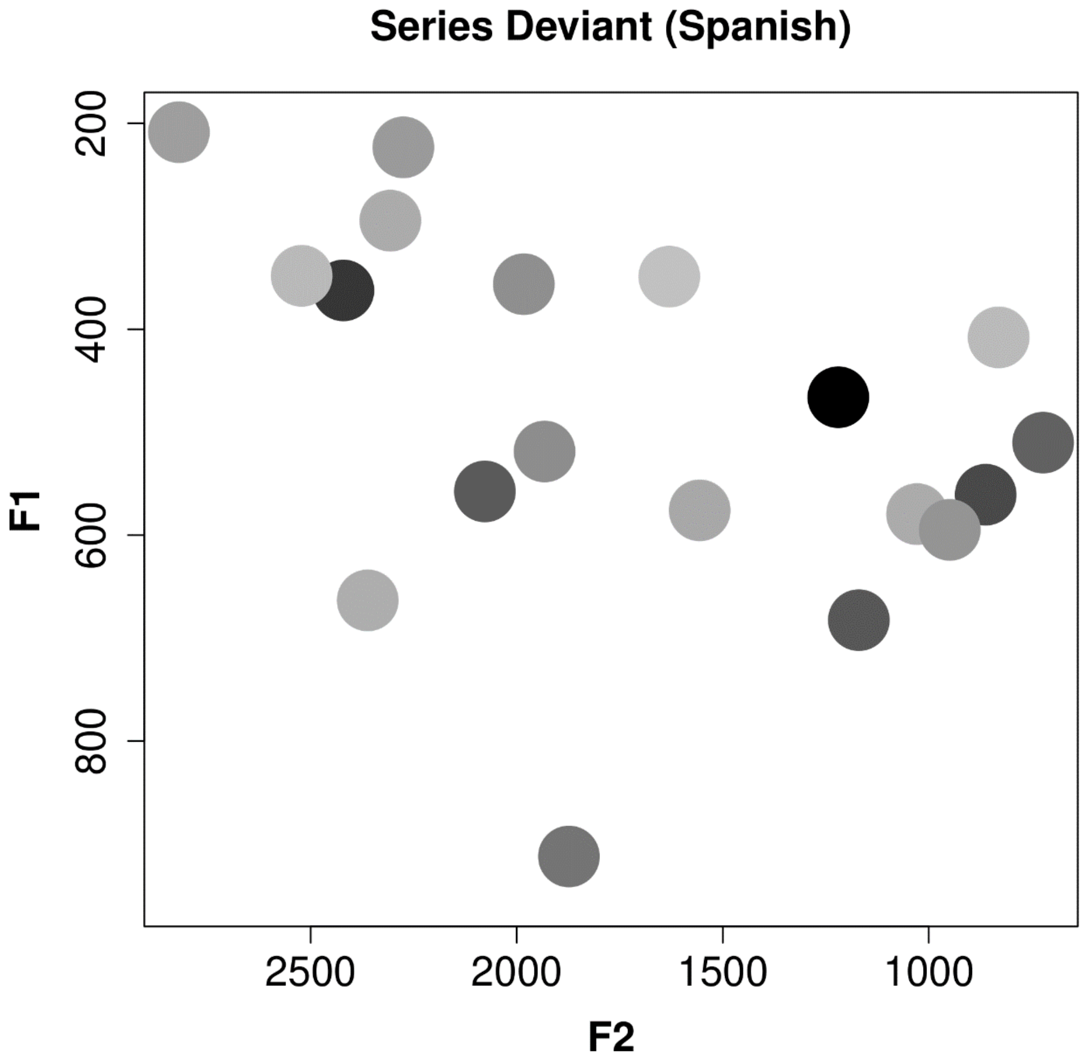

Figure 4.11. Vowel Plot of MMN Amplitudes for Series Deviants in Spanish Participants.

Figure 4.10. and Figure 4.11. illustrate different patterns of responses between the English and Spanish participants. Among the English participants, all high vowels elicited larger back-projected MMN amplitudes compared to the other IPA vowels. However, in the Spanish group, high vowels generally elicited smallest MMN amplitudes, with the exception of /u/. In the English participants, /ɑ/ also elicited large MMN amplitudes, followed by /œ/ and /ʌ/. In the Spanish participants, /u/ elicited the greatest amplitude followed by /ø/, /ɒ/, /ɛ/, /ɶ/ and /ɔ/. At first glance, this pattern of results does not seem to be influenced by the native phonemic repertoire of the listeners. To fully explore the set of

results, the pairwise deviant back-projected MMNs were plotted in Figure 4.12., for English participants, and Figure 4.13. for Spanish participants.

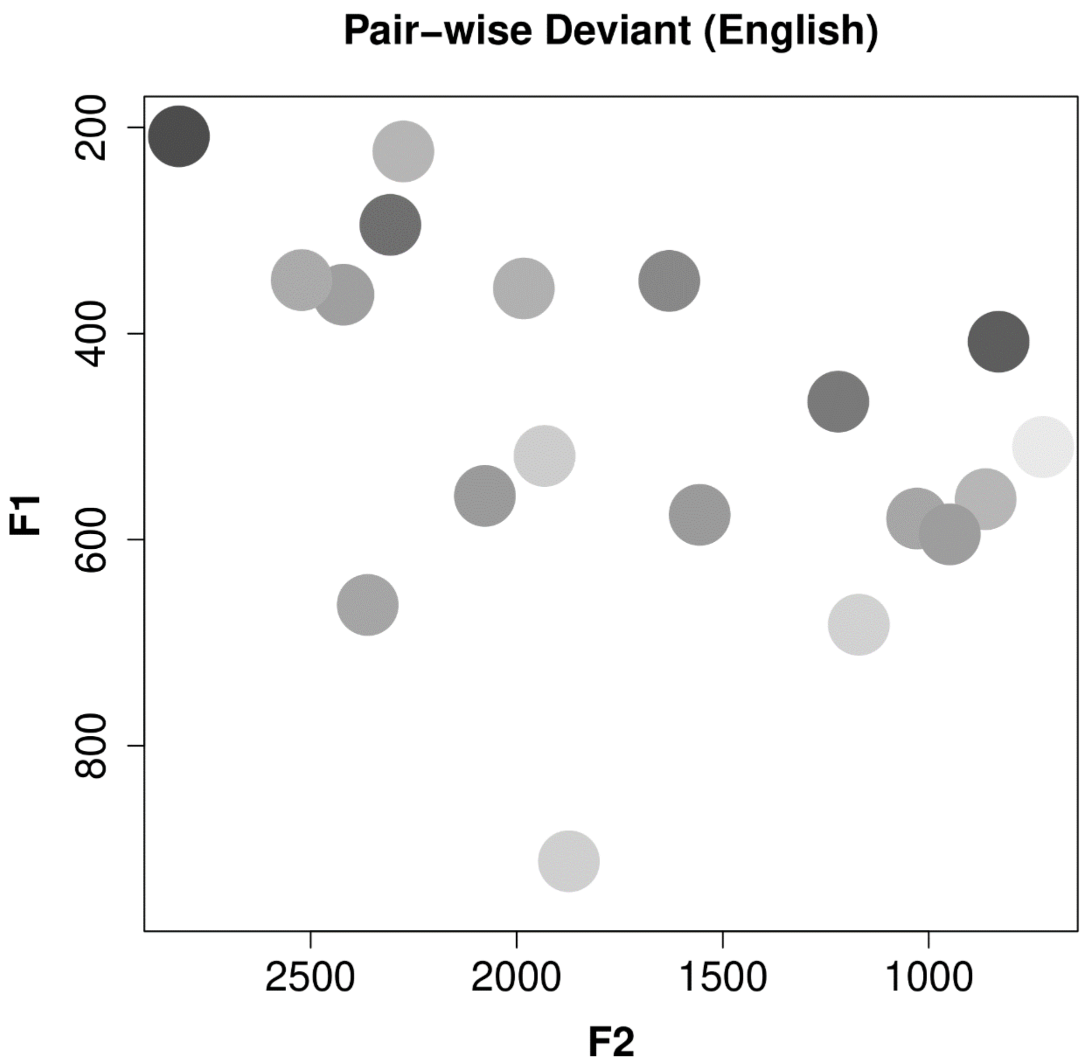

Figure 4.12. Vowel Plot of MMN Amplitudes for Pairwise Deviants in English Participants.

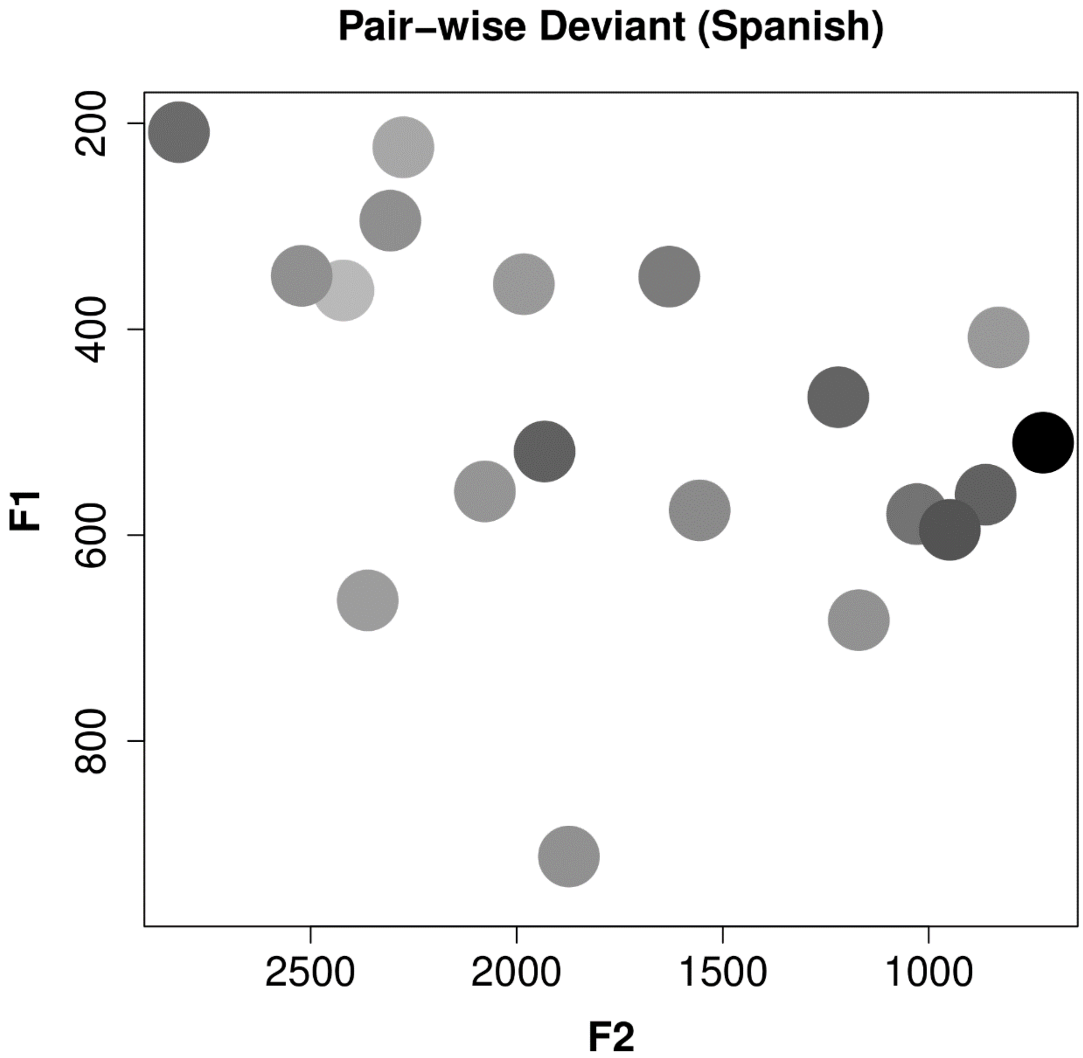

Figure 4.13. Vowel Plot of MMN Amplitudes for Pairwise Deviants in Spanish Participants.

Similarly to the series deviant results, English participants demonstrated larger MMNs to high and mid vowels, compared to lower vowels. In contrast, the Spanish group exhibited greater MMNs for mid-back and open-back vowels. Moreover, Figure 4.11. and Figure 4.13. demonstrate that the Spanish group show overall larger MMNs for the pairwise deviants than for the series deviants, as indicated by the statistically significant difference in MMN magnitudes between the series deviants and the pairwise deviants in the Spanish participants. In the pairwise deviant MMNs, the Spanish group demonstrated a pattern of results that corresponded loosely with their native phonemic repertoire. Specifically, the high front and back vowels /i/ and /u/, elicited relatively large MMNs. Large MMNs were observed in the mid-front region, although the vowel that elicited these responses was /œ/ rather than /e/, which is the closest to the native exemplar. Additionally, the mid-back vowels elicited large MMNs, although the largest MMN was attributed to /ɔ/ and not /o/, which more closely resembles the Spanish mid-back vowel. The central IPA vowels elicited smaller MMNs, potentially due to the absence of phonemic central vowels in Spanish, aside from /a/, which is the only central vowel in its system. Nonetheless, the open-front vowel /a/ did not elicit large MMNs in Spanish speakers. Therefore, although the Spanish speakers' pairwise MMNs reveal the pattern of results that most closely aligns with their native phonemic repertoire, the MMN magnitudes do not directly correspond to the native phonemes of these speakers.

4.3.4. Effects of Vowel Assimilation in MMN magnitudes

Since the identification scores on the behavioural task did not directly map onto the native phonemic repertoire in either speaker groups, we aimed to examine whether the amplitude of the back-projected MMNs was influenced by the degree of assimilation of each IPA vowel into a native vowel category. For this purpose, we built a mixed-effects model in which the predictor, which accounted for the variance across IPA vowels was the mean identification score of each IPA vowel, within each language group. These scores were obtained from the behavioural vowel identification tasks and were included in the model as a continuous fixed effect. The model also had a fixed effect of listener's native language (English vs. Spanish); a continuous fixed effect of the acoustic distance between the previous and the current stimulus heard and a categorical fixed effect of deviant type (series vs. pairwise). Additionally, by-subject random effects were included to account for individual variability. Once again, the model's dependent variable was the peak amplitude of the stimulus reconstruction, measured for each deviant stimulus and each subject. This means that single-trial deviant data was analysed for every subject and only the deviant responses were included in the statistical analyses.

The test of Analysis of Deviance using Type II Wald chi-square tests revealed a significant main effect of assimilation scores (χ^2 (1) = 9.89, p =0.002) and a significant main effect of deviant type (χ^2 (1) = 4.03, p =0.045). Moreover, the interaction between assimilation scores and deviant type was also significant (χ^2 (1) =7.74, p =0.005). No other main or interaction effects were significant. Since the language background of the subjects did not show a significant main effect on the MMN amplitudes, when accounting for the assimilation scores of the vowels, we compared the full model to a reduced model, which excluded the language background predictor. The full model had 10 parameters (AIC = -3785, BIC = -3704, log$L$ = 1902, D= -3805), while the reduced model had 8 parameters (AIC = -3788, BIC = -3724, log$L$ = 1902, D= -3805). The reduced model had lower AIC and BIC values, indicating a better fit. Moreover, an ANOVA comparing the two models showed that the addition of the language background predictor did not significantly improve the model fit (χ^2(2) = 0.07, p = 0.967). Therefore, we excluded the language background predictor.

To further simplify the model; in the next analysis step, we compared the reduced model to a simpler model that excluded the acoustic distance predictor, as it had not shown main or interaction effects in the analysis of deviance tests. The simpler model (k=6, AIC=-3789, BIC=-3741, log$L$=1901, D= -3912) had lower AIC and BIC values than the reduced model, indicating a better fit. Additionally, an ANOVA comparing the two models showed that the inclusion of the acoustic distance predictor did not significantly improve the model fit (χ^2(2) = 2.87, p = 0.24). Therefore, the simpler model, which excluded the language background and the acoustic distance predictors, was selected as the final model.

Tests of Analysis of Deviance on the final model showed a significant main effect of assimilation scores (χ^2 (1) = 8.91, *p* =0.003). The results of a Pearson's r correlation analysis indicated that those IPA vowels with the highest degree of assimilation, generated MMNs of greater amplitude r (23864) = 0.025, p = 0.001. The 95% confidence interval was [0.012, 0.037]. Additionally, there was a significant main effect of deviant type (χ^2 (1) = 4.05, *p* =0.044) on the back-projected MMNs and the interaction between assimilation scores and deviant type was significant (χ^2 (1) =7.77, *p* =0.005). Meaning that, the effects of assimilation scores on the MMN amplitude were influenced by the type of deviant that subjects heard. Post hoc analyses were conducted using the *emmeans* package to further investigate this interaction. Pairwise comparisons revealed that the effect of assimilation scores on MMN amplitude differed significantly between deviant types (p = 0.0443, SE = 0.00289, z = -2.012). The estimated marginal means for MMN amplitude were calculated for each combination of assimilation scores and deviant type. These results showed that, at an assimilation score of 0.646, pairwise deviants elicited larger MMNs (EMM=0.024) than the series deviants (EMM=0.018). Figure 4.14. and Figure 4.15. present a graphical representation of the results in the English and the Spanish groups, respectively.

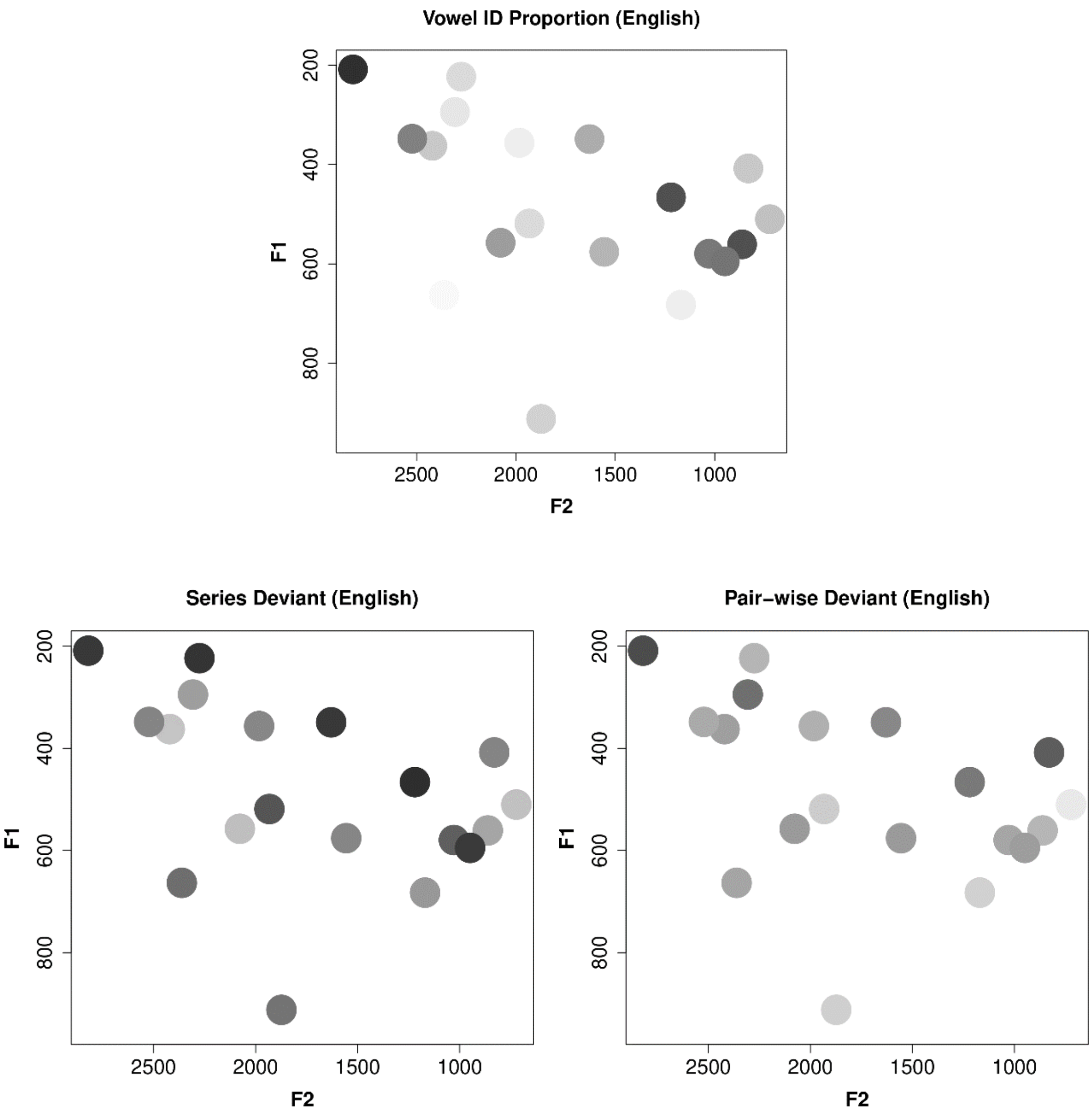

Figure 4.14. Vowel Identification and MMN Plots for English Participants. The top graph shows the vowel identification scores. Each dot represents an IPA vowel and the darkness of each dot indicates the degree of assimilation of the vowel, into a native vowel category. The darker the dot, the greater the assimilation into a native category. The series and pairwise deviant MMN plots are shown side by side.

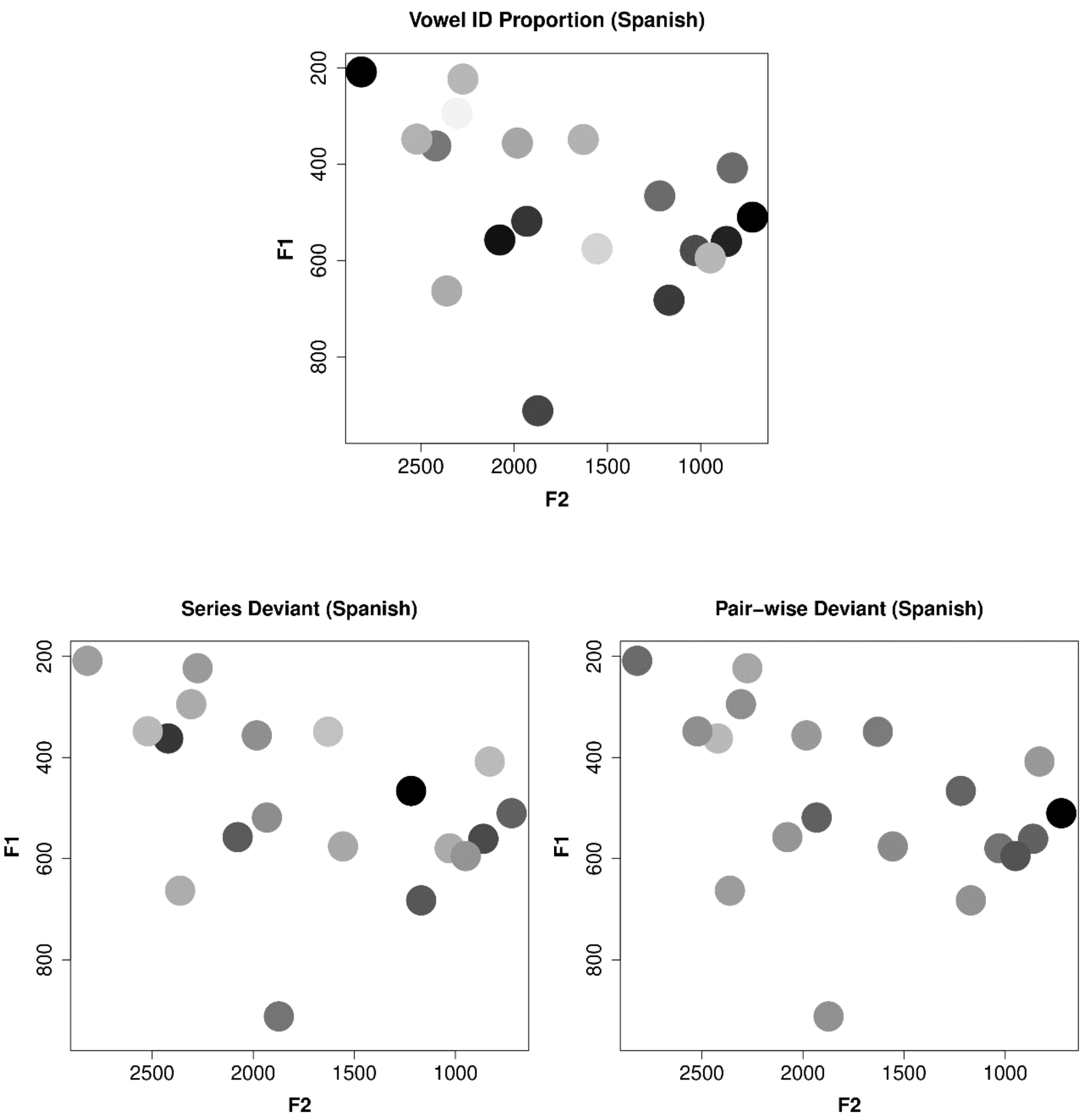

Figure 4.15. Vowel Identification and MMN Plots for Spanish Participants. The top graph shows the vowel identification scores. Each dot represents an IPA vowel and the darkness of each dot indicates the degree of assimilation of the vowel, into a native vowel category. The darker the dot, the greater the assimilation into a native category. The series and pairwise deviant MMN plots are shown side by side.

Although the statistical analyses revealed a significant effect of vowel assimilation on the amplitude of the back-projected MMNs, the results in Figure 4.14 and Figure 4.15 do not demonstrate a straightforward mapping between the MMN amplitudes and the identification scores, in either deviant type. The Spanish pairwise results display the pattern of MMN amplitudes that most closely aligns with the identification scores of each vowel: the high vowels /i/ and /u/ demonstrated large and moderate assimilations, respectively, and they showed moderate pairwise deviant MMNs. Moreover, the cluster of mid and open back vowels elicited moderate to large assimilation scores and moderate to large pairwise deviant MMNs. However, the mid-front vowel was strongly assimilated into a native category but generated small MMNs and the same pattern was seen in the open front vowel.

4.3.5. The Perceptual Maps

Since the results demonstrated effects of vowel assimilation on the elicited MMNs, we carried out an additional data visualisation step in which we created vowel maps for both participant groups. Each IPA vowel was assigned vowel height and front/back values to create maps that resemble a traditional vowel quadrilateral. The vowels were once again, represented with dots which vary in darkness, with darker dots indicating larger back-projected MMNs.

Figure 4.16. and Figure 4.17. display the vowel maps as shown in the series and the pairwise deviants for the English and the Spanish groups, respectively.

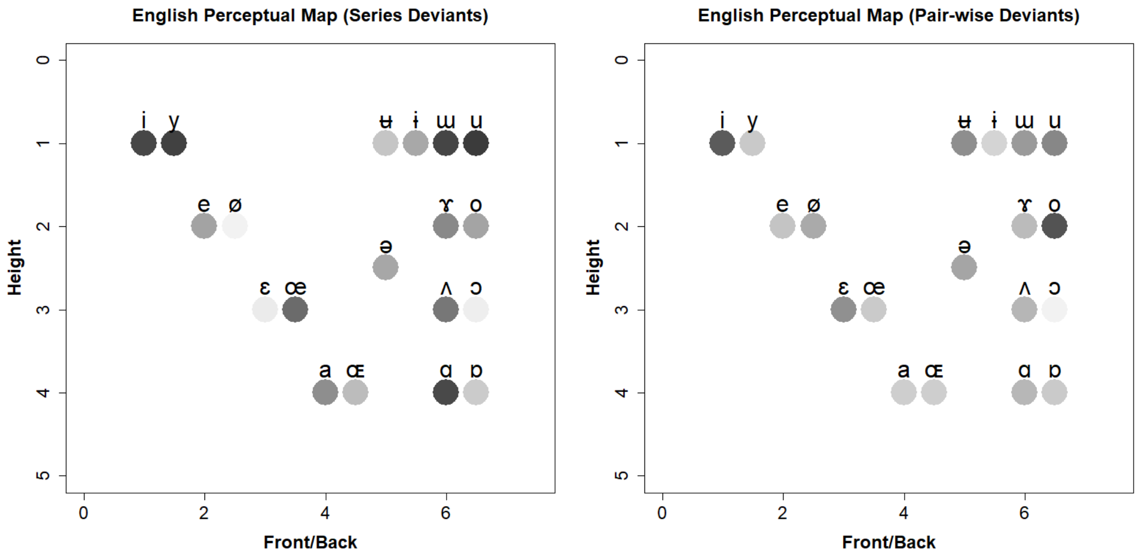

Figure 4.16. Perceptual Maps for English Speakers. The 19 IPA vowel stimuli are represented in a vowel quadrilateral to depict their articulatory features. Each vowel is labelled with its corresponding IPA symbol. The darker the dot representing the vowel, the greater the averaged back-projected MMN that it elicited. Series deviant results are plotted to the left and pairwise deviant results, to the right.

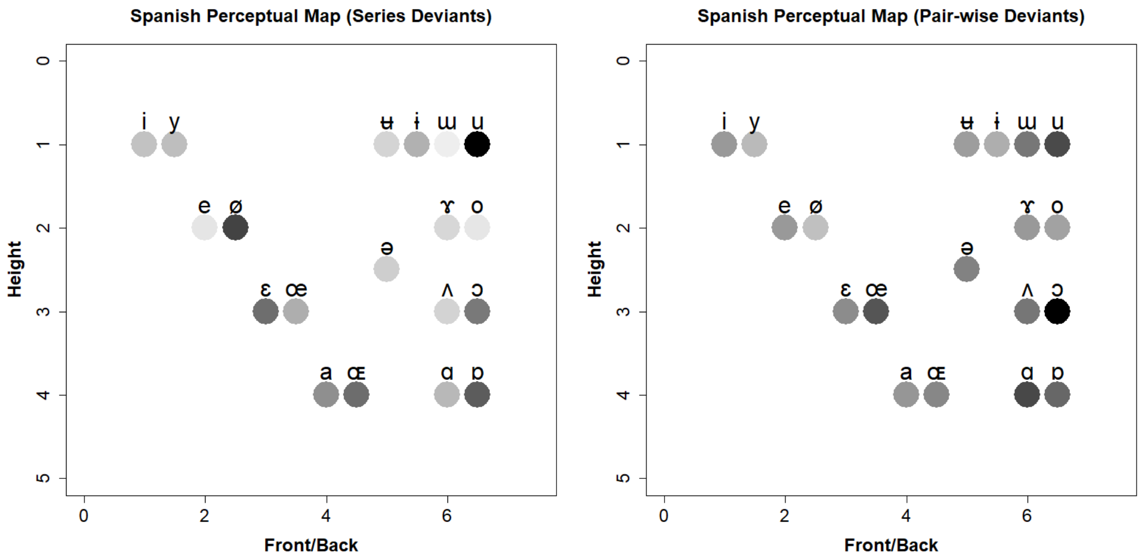

Figure 4.17. Perceptual Maps for Spanish speakers. Each of the 19 IPA vowel stimuli is shown in a vowel quadrilateral according to its articulatory features, labelled with its IPA symbol. Darker dots indicate greater averaged back-projected MMN amplitudes. Series deviant data is shown on the left; pairwise deviant data on the right.

The perceptual maps generated using the back-projected MMNs, demonstrated different patterns of MMN sensitivity across both speaker groups. The series deviant maps demonstrate a pattern of results that could be rooted in the perception of focal vowels since the vowels at the extremes of the quadrilateral elicited the largest MMNs. Once again, the map that most closely maps onto the native phonemic repertoire of the listeners is the Spanish pairwise deviant map. In this map, the native-like /i/, /e/ and /u/ demonstrate moderate MMNs. However, /o/ and /a/ would be expected to generate larger MMNs since they are also native Spanish vowels.

4.4. Discussion

The current study used a vowel-pair oddball paradigm with 19 IPA vowels to map out the perceptual vowel spaces of Southern British English speakers and Spanish Speakers. The current oddball paradigm was a reversed version of the paradigm used in chapter 3 in which the deviant stimuli, in both deviant types, entailed a phonetic difference between the expected vowel and the actual vowel that subjects heard. The amplitude of the recorded MMNs was influenced by the effect that the language background of the listener had on each IPA vowel deviant. Moreover, the degree of assimilation of each IPA vowel into a native vowel category, predicted the amplitude of the MMN that each vowel generated - the vowels which were perceived as better exemplars of a native category, elicited MMNs of greater amplitude. Therefore, the MMNs recorded in the current experiment demonstrate a clear phonetic component. Meaning that our MMN framework is a useful tool for exploring cross-language vowel perception.

The perceptual maps in Figure 4.16 and Figure 4.17 show a complex pattern of results, which do not correspond directly to the native repertoires of either language groups. The IPA vowels, which are most similar to the native exemplars, were expected to elicit larger MMNs (Näätänen et al., 2007a). However, the English participants showed increased series-deviant MMNs for most high vowels, including non-native ones, and overall small pairwise deviant MMNs, even for native-like vowels. This pattern of results could reflect the low assimilation scores that the IPA vowels elicited in the English- speaking group. Nonetheless, the Spanish speakers, who demonstrated more reliable assimilation of the IPA vowels into native categories, do not show MMN sensitivity to native-like vowels in the series deviant results (i.e., the only native vowel with large MMNs is /u/ and the second largest MMN magnitudes were elicited by the non-native vowels /ø/ and /ɒ/). The Spanish pairwise deviant data revealed the pattern of MMN magnitudes that most closely translated to the listeners' native phonemic repertoire, but non-native vowels like /ɔ/ and / œ/ still generated large MMNs compared to the rest of vowels.

A potential explanation for this lack of MMN sensitivity to native contrasts could lie in the nature of the experimental stimuli. Previous investigations that have reported reliably greater MMNs to native phonemes, compared to non-native ones, utilised sets of stimuli which differed in small steps along a specific acoustic dimension (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Sharma & Dorman, 1998). These findings could also be influenced by subject's perceptual biases towards prototypical exemplars of a vowel category (see Ylinen et al., 2010). That is, MMNs to prototypical phonemes, as compared to acoustically similar phonemes which are less prototypical examples, will be larger because these prototypical phonemes are likely perceived as more salient. The current stimulus set expanded the whole IPA vowel space and did not contain prototypical vowel examples for either language group. Moreover, it has been suggested that the predictive model generating the MMN is based on the standard stimuli. The standard vowel pairs did not contain prototypical vowels either and this could have affected the MMN amplitude markedly. As a result, the vowels that were perceived as more salient and hence, generated larger MMNs were not always those which resembled native vowel exemplars.

Importantly, scores on the vowel identification and category goodness task did also not fully align with the native vowel repertoires, especially in the English participants. As a result, we sought to investigate how the degree of assimilation would influence the generated MMNs. The degree of assimilation into a native vowel category, directly predicted the MMN magnitude for each vowel. Despite this, Figure 4.16 and Figure 4.17 illustrate a complex pattern of results. Among the English participants, neither deviant type demonstrated a direct correlation between the assimilation scores and the MMN amplitudes. The series deviant results exhibited greater contrasts in MMN amplitudes, with some vowels showing moderate to large MMNs, while others displayed very small MMNs. In contrast, the Spanish participants' series deviants generated relatively low MMNs compared to the magnitude of their assimilation scores, and the pairwise deviant MMNs aligned more closely with the assimilation scores, although this alignment was not perfect. Unfortunately, the current study lacked sufficient statistical power to analyse each IPA contrast individually, which may explain why the *emmeans* analyses did not yield statistically significant results for specific contrasts. Nonetheless, these results show that the current MMN framework can illustrate distinct patterns of vowel perception amongst different listener groups.

A primary aim of the current investigation was to ensure that our MMN tool could be applied to different languages, rather than being limited to Spanish and English participants. Therefore, we chose IPA vowels over native-like vowel exemplars, as they are a universal set of phonemes. Despite this, the Spanish speakers demonstrated stronger assimilation of the IPA vowels into native categories. In the English group, the identification responses showed less consistent mappings between IPA vowels and native vowel categories, potentially because English speakers have a larger native vowel repertoire. According to the perceptual magnet model (Iverson & Kuhl, 1995; Kuhl, 1991), native phonetic categories induce prototypes of the native vowels, which act as perceptual magnets, drawing similar vowel sounds towards the ideal exemplar in the native language. The five native vowel categories of Spanish speakers could have served as strong perceptual magnets for the non-native IPA vowels. In contrast, native English speakers, who are accustomed to tuning into subtle acoustic vowel differences might have faced a greater challenge in categorising the unfamiliar IPA vowels. As a result, they demonstrated low identification scores and a relatively unexpected pattern of vowel assimilation. Importantly, although the number of vowels in each language could be seen as a confounding factor in our study, we chose to not control for it because, in their current form our results accurately reflect the inherent characteristics of both languages. That is, it is possible that speakers of languages with fewer vowels might be better suited for this type of MMN investigation.

Nonetheless, the MMNs recorded in both speaker groups seem to have a phonetic component, as opposed to the MMNs recorded in chapter 3, which mostly reflected effects of perceptual biases favouring focal vowels i.e., those with prominent spectral peaks. In the current investigation, the English series deviant results, demonstrated larger MMNs to vowels at the extremes of the vowel space, which could be rooted in this perceptual bias. However, the methodological modifications implemented in this study allowed us to record different patterns of MMN amplitudes across the speaker groups, reflecting the differences in their native vowel repertoires. In the current paradigm, the deviant stimuli consisted of IPA vowels, which differed acoustically and phonetically from the standard stimuli. This methodological modification goes in line with the majority of previous work on phonetic contributions of the MMN, which used deviants belonging to different phonemic categories to the standard stimuli (reviewed in Näätänen et al., 2007).

Another methodological difference between the two studies lies in the stimulus function used for the mTRF analyses. In Chapter 3, we designed a stimulus function specifically tailored to capture distinctions between standard and deviant responses, which was used as the acoustic input representation in the mTRF model. In contrast, the current analysis did not encode the differences between standards and deviants categorically. Instead, it modelled the acoustic encoding of the vowel pairs over time. The data from the 64 EEG channels were back-projected onto an amplitude envelope function representing a vowel pair. Notably, these back-projected responses manifested as increased positivity for deviant stimuli, rather than the typical negativity commonly associated with MMN responses. Moreover, the MMNs were not realised as smaller neural responses but instead, reflected enhanced tracking of the amplitude envelope in the deviant stimuli. Figure 4.8 and Figure 4.9 revealed differential neural responses between standards and deviants with series deviants displaying the increased positivity in the first vowel and pairwise deviants showing increased positivity in the second vowel, as compared to the standards. In both cases, the increased positivity corresponded to the stimulus that introduced the deviance.

Further analysis of the back-projected data revealed notable differences in the overall MMN amplitude across both deviant types. Pairwise deviants elicited MMNs of a significantly greater amplitude than the series deviants. This larger response for pairwise deviants responses could be attributed to our averaging method: pairwise deviant responses were averaged over a shorter 200ms window from the onset of the second vowel, compared to a 400ms window from the onset of the first vowel for series deviants. The averaging windows were chosen based on the timing of the introduced deviance. Additionally, the deviance induced by the first vowel in the series deviants may not have appeared as salient to the listeners, as the deviance induced by the different second vowel in the pairwise deviants. It is possible that participants habituated to the alternating listening condition, where the standard vowel pair changed continuously, while the introduction of a deviant second vowel in the pair continued to elicit a salient mismatch response. Moreover, only the pairwise deviant responses showed an interaction with the degree of vowel assimilation in the Spanish group. This indicates that the effects of assimilation scores on MMN amplitude are most pronounced under conditions of higher overall assimilation scores, which correspond to MMNs of greater amplitude. In other words, the relationship between vowel assimilation and neural responses is strongest when the stimuli exhibit more robust assimilation patterns, leading to larger MMN amplitudes.

To conclude, the current investigation has demonstrated that the reported MMN tool is optimal for exploring cross-language vowel perception. Our results show reliable evidence that the recorded MMNs could serve as neural indices of vowel perception since a) they show different patterns of amplitude across English and Spanish speakers and b) the degree of assimilation of the deviant vowels influenced the amplitude of the generated MMNs. Nonetheless, the perceptual maps created using the back-projected MMN amplitudes did not fully correspond to the native phonemic repertoires of either speaker group. The choice of IPA stimuli could have impacted these findings since the stimulus set varied across several acoustic dimensions, inhibiting the MMN's reported sensitivity to native/native-like contrasts. The effectiveness of the tool may also be impacted by the size of the listeners' native vowel repertoire, since here, the speakers with the larger vowel inventory demonstrated inconsistent patterns of assimilation of IPA vowels. Moreover, the relationship between vowel assimilation and MMN amplitude was strongest when the stimuli were more strongly assimilated into native vowel categories. Overall, this tool needs further refining and the current investigation presents an advancement from the findings of Chapter 3 and highlights the tool's potential to record neural indices of perception of several phonetic contrasts across different speaker groups.

5. General Discussion

The current thesis presented two tasks for measuring phonetic perception, both behaviourally and at the neural level. The overall aim was to design protocols for testing phonetic perceptual abilities that would allow for time-effective data collection. In particular, the current thesis explored phonetic learning in a high variability phonetic training (HVPT) program with primary-school children and an electroencephalography (EEG) technique comprising a vowel-pair oddball paradigm for Mismatch Negativity (MMN) response elicitation and a machine learning analysis methodology. Together, these two techniques allowed for the collection of perceptual data from over one hundred children and provided a means of recording adult neural indices of perception to 19 vowel contrasts in a single study. Therefore, both frameworks proved successful in time-effective data collection. This general discussion will explain in detail the main findings of the current thesis and will suggest ways in which both, the behavioural and the electrophysiological methodologies described in this thesis could be used in combination in future studies investigating phonetic learning.

In short, there are three main contributions to the literature resulting from the current thesis. Firstly, the HVPT program demonstrated that older Spanish-Catalan children who used both of their languages most often were the best improvers in their perception of English vowel perception. These findings add to the literature on age and bilingualism effects in training and have important implications for second language in-classroom instruction. Secondly, Chapters 3 and 4 of this thesis demonstrated that multivariate Temporal Response Functions (mTRFs; Crosse et al., 2016) are well-suited for MMN data analysis. This alternative framework of MMN testing offers an alternative to the standard combination of single-contrast oddball paradigms and event related potential (ERP) analysis of the EEG data. Thirdly, the mTRF analysis methodology for MMN analysis reveals cross-language effects in IPA vowel perception across Spanish and English speakers. Therefore, they provide an adequate means of exploring cross-language vowel perception at the neural level.

Prior to the HVPT study reported in the current thesis, the memory-card game training technique was tested with Japanese adults and a smaller sample of Spanish-Catalan children (Iverson et al., 2023). In this pilot investigation, the forced-choice oddity task was incorporated in every testing and training session and it had fewer animated characters speaking words to the children. Results of the pilot, showed that children's discrimination performance of English vowels became worse over time. That is, they took longer to complete the trials and their accuracy in detecting the odd word, decreased. These results were considered to reflect effects of boredom, since besides the statistically tangible decrease in performance accuracy, children reported wanting to play with more animated animals. As a result, the oddity task was modified to include three more animated characters and the training schedule was revised to include the oddity task only on the pre- and post-test sessions, leaving the training sessions to only include the memory-card game. Since the HVPT methodology had already been streamlined for child testing, the first study of the thesis was able to focus on answering open-ended questions in the child phonetic training literature.

In contrast, Chapter 3 reports an initial investigation of the proposed MMN framework. Previous MMN studies of phonological processing had included few phonetic contrasts and had mostly used an ERP analysis (Chládková et al., 2013; Dehaene-Lambertz, 1997; Hisagi et al., 2010; Kazanina et al., 2006; Näätänen et al., 1997; Sharma & Dorman, 1999, 2000), with the exception of investigations using single-trial MMN data analysis techniques (Brandmeyer et al., 2013). Therefore, the investigation in Chapter 3 was purely exploratory and had a greater focus on the development of the experimental methodology. Results showed that the vowel-pair oddball paradigm elicited MMN responses, with the caveat that they did not show clear influences of vowel categorization. Following the methodological evaluation, Chapter 4 modified the oddball paradigm and included a variation of the mTRF analysis to apply the MMN technique to a larger sample size and explore cross-language vowel perception. Consequently, the experiment in Chapter 4 was able to include theoretical considerations of cross-language perceptual abilities as well as methodological considerations of MMN elicitation and data analysis.

5.1. Age and bilingualism effects in phonetic learning

To date, there is a small but growing body of literature on child HVPT demonstrating that children's phonological advantage in second language (L2) learning, over adults, does not always translate in a better performance in HVPT investigations (Heeren & Schouten, 2008, 2010; Shinohara & Iverson, 2021; Wang & Kuhl, 2003). Moreover, in these training protocols, older children tend to outperform younger children (Brekelmans, 2020; Giannakopoulou et al., 2013; Kasisopa et al., 2018; Shinohara & Iverson, 2013) and teenagers outperform older children (Shinohara & Iverson, 2021). In the HVPT study of the current thesis the training paradigm consisted of a memory card game in which children had to remember the locations of cards, placed face-down on a grid, to find the pairs. Each card was linked to a word and pairs consisted of two cards linked to the same word, spoken by two different speakers. In some blocks of the game, the cards revealed a symbol on their flip side when they were turned over and children could combine the audio and visual information to find the matches. This training paradigm differs greatly from conventional HVPT training using identification and discrimination tasks with trial-by-trial feedback.

Firstly, in phonetic identification and discrimination tasks such as those used by Brekelmans, (2020), Kasisopa et al., (2018) and Shinohara and Iverson (2013), participants' aim is to improve their phoneme perception, rather than to perform well in a game. As a result, they are likely to focus their attention on specific acoustic cues that enable them to distinguish the phonetic contrasts on which they are being trained. Secondly, trial-by-trial feedback can help subjects learn the rules that distinguish the target phonemes, and consequently, participants are able to consciously apply those rules whilst completing the training tasks, as well as the pre and post test tasks. Lastly, in discrimination tasks, where subjects must judge whether an incoming phoneme is the same or different to the previous one(s), subjects must hold accurate representations of the phonemes in their memory for comparison. As a result, identification and discrimination tasks engage cognitive abilities that are more developed in older children, resulting in better performance than younger children.

In contrast, in the memory card game, children must associate sounds with images, remember card locations, and switch between visually searching the grid and, in some blocks, processing the words acoustically. In the game, exposure to high variability phonetic input is incidental and players are not given overt trial-by-trial feedback to direct their learning. Meaning that although the memory card game engages the visuospatial working memory to a greater extent than conventional identification and discrimination tasks, are not required to develop precise phonetic representations of the auditory stimuli. Moreover, in the blocks that show very distinctive symbols on the flip side of the cards (i.e., the easy symbols condition), children do not need to use the auditory stimuli to perform well in the game. Nevertheless, although our training paradigm had lower cognitive demands, older children demonstrated better performance than the younger children. A potential reason for this could be the nature of the pre and post test task.

The three-alternative forced-choice oddity task used to track phonetic perception from pre to post test created an explicit learning environment. Children were instructed to identify the odd word and hence, they understood the task to have a language focus. As a result their attention was directed to the acoustic details that differentiated the words in the task. Moreover, they received trial-by-trial feedback and had the chance to adjust their strategies depending on their performance. The multiple variations of phoneme discrimination tasks used in past HVPT studies differ in their level of complexity, and among them, the oddity task is the most demanding (Hansen Edwards & Zampini, 2008). This paradigm requires participants to hold multiple auditory inputs in memory and make comparisons under conditions of high uncertainty, both in terms of which sounds will appear and their order within the trial. As the complexity of the task increases, performance depends not only on the ability to perceive fine-grained sound differences, but also on more advanced cognitive skills such as the ability to retain an accurate representation of a previously heard sound whilst evaluating whether the incoming sound is the same or different. Therefore, as mentioned above, their increased working memory abilities, provided older children with an advantage in performance in the oddity discrimination task (Snowling et al., 1997).

It is also possible that the younger children in the current study might have shown similar levels of improvement in the oddity task, to those of the older children, if they had received more training sessions. The present HVPT intervention was relatively short due to constraints imposed by other school activities. In their review of second language learning, Krashen et al., (1979) noted that the learning advantage of older children tends to be limited to the initial stages of language acquisition. DeKeyser and Larson-Hall (2005) similarly argued that this early advantage stems from older children's greater ability to engage in and apply faster, more explicit learning processes. In contrast, younger learners rely more heavily on implicit learning mechanisms, which may explain their slower initial progress. Muñoz (2006) conducted a longitudinal study on second language acquisition across multiple linguistic domains. Their findings indicated that age differences at the onset of language learning tend to favour older learners in the short term, largely due to their more advanced cognitive development and the continued maturation of explicit learning systems. In contexts where opportunities for implicit learning and naturalistic practice are limited, older learners may acquire language features that depend on declarative knowledge and explicit memory more efficiently. Younger learners, by contrast, may be disadvantaged in such environments, not because of lower learning potential, but because limited exposure hampers the implicit, input-driven learning processes that typically underpin first language acquisition.

Moreover, the current thesis demonstrates that bilinguals have an advantage in HVPT over monolinguals, in the later years of childhood. This advantage can be explained in terms of the breadth of their phonological repertoire and, their metalinguistic knowledge. It is well-known that in HVPT, participants apply their pre-existing phonetic categories to process the auditory stimuli. The vowel inventories of Spanish and Majorcan Catalan differ in that, besides the 5 vowel categories present in Spanish (i.e., /i/,/e/, /a/, /o/and, /u/), Majorcan Catalan has three additional categories in the mid region (i.e., open-mid front /ɛ/, open-mid back /ɔ/ and the central vowel /ə/). Therefore, children who use Spanish and Catalan more often, have more native vowel categories. As they acquire linguistic experience, with age, these native categories become more robust and older bilingual children are able to leverage their knowledge of native categories to learn English vowels. These findings go in line with those in Iverson and Evans (2009), showing that listeners with larger vowel repertoires, (i.e., German speakers), were able to learn non-native vowels quicker than listeners with fewer native vowel categories, (i.e., Spanish speakers). Moreover, the enhancing effects of a larger vowel repertoire may have interacted with the effects of increased metalinguistic awareness in the bilinguals. Children who grow up bilingual often develop stronger metalinguistic skills, allowing them to recognize and think critically about language structures (Jensen, 2008). This heightened sensitivity may also mean they are more familiar with the idea that each language has its own unique sound system. As a result, they could be better at detecting unfamiliar phonemes. In our sample, older children have likely spent more time studying English and may have developed more effective learning techniques. Therefore, those from bilingual backgrounds might be particularly skilled at picking up on relevant cues in the training materials.

Importantly, the literature demonstrates that, from a young age, bilinguals are more susceptible to the effects of task demands than monolinguals. For instance, when measured with a head-turn preference procedure, bilingual infants' ability to discriminate native vowel contrasts shows a developmental delay, compared to monolingual performance (Bosch & Sebastian-Galles, 2003). However, when an anticipatory eye movement paradigm, is used instead, bilinguals do not show any delays in perceptual abilities (Albareda-Castellot et al., 2011). Perhaps these results reflect that bilinguals' ability to discriminate between vowel sounds is not as robust as that of monolinguals. The results from the current HVPT study indicate that in early childhood, bilinguals may still have less robust discrimination abilities since the younger bilinguals in our study started off with lower discrimination accuracy in the three-alternative oddity task.

Nevertheless, Chapter 2 shows that, as bilinguals gain experience and linguistic knowledge in both of their languages, they become better learners of English vowels than their monolingual peers. This bilingual advantage in phonetic/phonological learning has also been demonstrated in children (Kasisopa et al., 2018) and adults (Antoniou et al., 2015) for whom one of their languages shares similarities with the target language. Additionally, this advantage has been shown to extend to bilingual adults trained on phonetic contrasts absent in both of their native languages (Tremblay & Sabourin, 2012). All together, these findings demonstrate that being proficient in more than one language is beneficial when acquiring the sound system of an additional language.

5.2. In-classroom L2 learning

Moreover, the study on Chapter 2 could be used to draw conclusions on the outcomes of in-classroom L2 speech learning in early childhood. In Chapter 2, 6 year olds were the worse performers and the literature highlights a disadvantage in non-native phoneme discrimination at 4 and at 6 years of age. Werker and Tees, (1983) demonstrated that, when discriminating non-native contrasts of a language that children are not trying to learn, 4 year olds perform worse than 8 and 12 year olds. Burnham, (2003) positioned this decrease in non-native discrimination at 6 years of age, instead of at 4, and argued that it reflected a second period of perceptual attunement caused by the start of reading instruction in the native language. There are two main methodological differences that could have allowed Burnham (2003) to demonstrate better non-native discrimination in 4-year olds, than in 6-year olds: the type of task and stimuli used to measure children's discrimination performance. Werker and Tees used an AX discrimination task and Burnham used a conditioned head-turn preference task, which poses a lower cognitive load, especially on the younger children. Moreover, Burnham used a non-native voicing contrast (/b/-/p/) whereas Werker and Tees's contrast (/t̪a/-/ʈa/) involved a difference in the place of articulation, which is not present in the native phonology. Considering that later studies contributed supporting evidence to Burnham's (2003) hypothesis (e.g., Kasisopa et al., 2018), it is possible that adapting the task demands to the children's perceptual abilities could have allowed Burnham (2003) to accurately identify a particular difficulty in 6 year olds in discriminating non-native sounds.

Having said this, the children in Chapter 2 were tested in a language that they were learning. The literature on L2 learning in childhood demonstrates that 4 year olds who acquire an L2 after they have learned their L1, can acquire native-like perceptual abilities in their L2, if they receive L2 input during school hours and outside school (McCarthy et al., 2014). However, if their in-classroom, as well as their home linguistic input is divided between both of their languages, they will show worse discrimination abilities in one of their languages than their monolingual peers of that language (Darcy & Krüger, 2012). In Chapter 2, the children were learning English as an L2 but the amount of English input they received was constrained to two hours every week (i.e., the amount of time allocated to English instruction in the curriculum). It is possible that without large amounts of L2 input, the reported decrease in non-native perceptual abilities impacts children's English speech learning in the younger years of childhood, resulting in worse performance in HVPT in the 6-year olds.

Lastly, the finding from Chapter 2 could have practical implications on in-classroom L2 teaching. The HVPT study was conducted in the classroom in two different schools on the island of Majorca, Spain. In the Spanish national curriculum, English is taught as a foreign language from the age of three years. At preschool and in the younger years of primary school, oral tasks such as reciting the days of the week or singing songs are given priority over written tasks because children are still learning to read and write in their native language (L1) and adding orthography instruction in an L2 would pose a great challenge. As children become more competent readers and writers in their L1, they begin to engage in written communication in English. This causes a shift in the type of activities that children take part in during their English classes, as they progress through their primary education. As a result, younger children engage with L2 in the form of oral communication while older children tend to engage more in written communication in their L2.

Nevertheless, L2 speech perception and pronunciation abilities are crucial for fluent communication in an L2 and hence, they should be an important consideration in L2 English education. Although HVPT paradigms have been used extensively in the laboratory, they are used less often in the classroom. The current HVPT study demonstrates that incorporating training in the classroom helps children learn L2 English vowels. Both, the scores on the oddity task and the performance on the memory card game indicate that the children engaged with the gamified trainer throughout the training program, suggesting that the computerised tasks introduced an alternative educational tool, which children seemed to enjoy. Moreover, although this paradigm did not include L2 production measures, good perceptual abilities are a pre-requisite for optimal L2 production (Melnik-Leroy et al., 2022). Additionally, improvements in L2 perception have been shown to have a knock-on effect on L2 production (Sakai & Moorman, 2018). Therefore, incorporating HVPT trainers in the classroom could have beneficial outcomes for children across primary school and help them improve speech-processing abilities in their L2.

5.3. MMN for cross-language vowel perception

Demonstrating cross-language perceptual differences using behavioural tasks is very straightforward. Many studies have delivered different variations of discrimination and identification tasks to measure perception of native and non-native phonetic contrasts. However, these behavioural tasks cannot explain when those perceptual differences occur (e.g., Borst et al., 2013). On the one hand, cross-language differences could arise at low levels of perception and involve purely acoustic or phonetic processing. On the other hand, they could involve lexical access or other top-down process happening following phoneme recognition. In contrast, EEG measures allow us to explore early stages of processing. For instance, EEG directly records electrical activity generated by the brain whilst participants engage in perceptual tasks, offering an immediate insight into neural processes. It also provides millisecond-level temporal resolution, allowing researchers to capture rapid neural processes as they unfold in real-time (Zhang et al., 2020). This is crucial for studying the early stages of cognitive processing, which occur within milliseconds after a stimulus is presented. Therefore, the current thesis designed an EEG tool to explore cross-language vowel perception amongst English and Spanish speakers.

The neural response that was chosen was the MMN because several studies have demonstrated that MMNs show increased magnitudes to native phonemes, as opposed to non-native ones (e.g., Näätänen et al., 1997). This finding demonstrated that besides acoustic processing, phonetic and phonological process contribute towards MMN generation. Therefore, the MMN response represents an adequate index of vowel perception. Although soon after its discovery the MMN was used quite extensively to study phonetic and phonological processing, in recent years, fewer MMN publications have had a linguistic focus. This is because it is very time-consuming to test several phonological contrasts in traditional MMN paradigms. MMNs are small responses and many stimuli are needed to ensure that they are reliable recorded. The current thesis shortened the length of testing multiple phonetic contrasts by including two modifications to the traditional way of testing MMN: it involved a vowel pair oddball paradigm and a machine learning analysis method. Instead of a specific phoneme, the standard stimulus of the paradigm in Chapters 3 and 4 was a recurring sound pattern, in this case, a vowel pair. This allowed us to alternate the vowels in the standard and the deviant pairs at different time points, allowing us to include 19 different IPA vowels in a single paradigm, which had not been done before.

Results from Chapters 3 and 4 demonstrated the efficacy of employing backward models of multivariate temporal response functions (mTRFs; Crosse et al., 2016) for investigating the MMN. Importantly, Chapter 3 demonstrated that using the mTRF method, we could achieve large levels of MMN reliability after a single deviant instance in our paradigm with 19 different deviant types and with just 7 standard responses per deviant response. This is comparable to the reliability obtained with 100 deviants per deviant type tested in traditional ERP analysis. Therefore, this finding is very remarkable since mTRFs provide an alternative perspective of MMN analysis, which shortens the length of testing by reducing the amount of deviant instances needed in the experiment.

These mTRFs were applied using two types of stimulus functions, with the continuous stimulus function proving more effective in capturing MMNs indicative of phonological processing. In Chapter 3, the stimulus function categorized neural responses as either standard or deviant. Data analysis revealed signs of phonological processing, evidenced by left-lateralization of the response in the sensor spaces, a characteristic typically associated with language-processing tasks, as opposed to mere acoustic processing. However, the results did not conclusively identify a phonological component of the MMN, as MMNs with larger magnitudes did not correspond to better assimilation responses. Conversely, Chapter 4 applied an analysis technique, which examined MMN responses in a continuous manner and revealed distinct cross-language effects. Specifically, English and Spanish participants exhibited differential patterns of MMN magnitudes elicited by the 19 IPA vowel stimuli.

Nevertheless, the exact implications of the MMN magnitudes recorded in the current studies remain unclear. Chapter 3 revealed that MMN magnitudes were influenced by the second spectral moment (SM2) and the second formant frequency (F2) of the IPA vowels, suggesting these MMNs were not entirely randomly generated. Despite our consistent aim to elicit MMNs indicative of phonological processing, Chapter 4 did not produce perfect vowel maps for either speaker group. Specifically, those IPA vowels which resembled native vowels, did not elicit the greatest MMNs. In previous studies reporting increased sensitivity to native vowels, fewer acoustic stimuli were examined and the recorded MMNs reflected perception of fewer acoustic dimensions. Meaning that our broader range of acoustic stimuli and dimensions may have diluted the sensitivity to native vowels. Additionally, the lack of correspondence between MMNs and assimilation may be attributed to the stimuli not being perfect exemplars of native vowels for either the Spanish or the English listeners, as the IPA represent a universal set of vowels. The use of a universal set of IPA vowels most likely influenced the MMN magnitudes as larger MMN magnitudes are most like elicited by prototypical examples of native vowels (Ylinen et al., 2010). Altogether, the findings suggest that although we are able to illustrate cross-language differences in MMN generation, eliciting MMNs indicating increased sensitivity for native vowels requires careful consideration of stimulus selection and acoustic dimensions. Perhaps, replacing the IPA vowels with recordings of native vowels by native speakers of either language would illustrate a pattern of MMN magnitudes aligning with the native vowel repertoires of both speaker groups.

Importantly, the MMN technique developed in this thesis offers a robust method to assess phonetic learning induced by gamified training protocols such as the memory card game. MMN is elicited during passive listening and does not require overt responses. Therefore, it is ideal for testing children whose performance in behavioural tasks may be influenced by attention or cognitive maturity. As such, it enables fair comparisons across age groups. Including MMN as a complement to behavioural tasks can provide insight into learning outcomes and underlying mechanisms that behavioural measures alone cannot fully capture. While behavioural tasks typically assess conscious, attention-driven aspects of learning, MMN reflects automatic, pre-attentive neural processing of auditory stimuli. In future experiments, we could compare MMN amplitudes to the target phonemes before and after training, providing an objective measurement of neural changes as a result of training. The memory card game targets the entire Southern British English vowel set, and we have developed a platform capable of testing multiple vowel contrasts. Combining this training tool with MMN recording will allow us to evaluate the extent of learning across the English vowel system in young learners of English.

Previous investigations have combined gamified phonetic training with MMN studies to evaluate neural changes to the trained phonemes. These studies have shown that MMNs can capture training-related changes in speech perception; for example, Zhang et al. (2009) observed increased neural sensitivity to the English /r-l/ contrast in Japanese adults after training, and Ylinen et al. (2010) reported enhanced pre-attentive processing of spectral cues in Finnish learners of the English /i/-/ɪ/ contrast, as indicated by post-training MMN responses. Moreover, Junttila et al., (2022) tested the effects of gamified phonetic training on children's MMNs and found similar results. They recorded MMN magnitudes to the

English voiced and voiceless dental fricatives /θ/ and /ð/ in 7 to 11 year old Finnish-native children, both before and after receiving HVPT training. Their results showed that children who completed a gamified training program, which required them to imitate English words containing the target sounds, exhibited a greater increase in MMN amplitudes from pre- to post-test, suggesting enhanced neural sensitivity to the trained contrasts.

In light of these findings, I believe that with additional training sessions using the memory card game, children would show enhanced MMN responses to the British English vowel set – particularly for those phonetic contrasts that were easier to learn. Replacing the oddity discrimination task in Chapter 2 with an EEG experiment using an MMN paradigm would allow us to design a high-variability phonetic training protocol in which both the training and testing tasks reflect implicit learning. In the MMN setup, as in the adult experiments described in Chapters 3 and 4, children's attention would be directed toward silent cartoons they enjoy, rather than the auditory stimuli in the oddball sequence. This would remove the need for active language processing, which would differ greatly from the more explicit learning context of the oddity task used in Chapter 2. A key consideration for this future study would be that the training materials in the memory card game encompass a wide range of phonetic contrasts. It may be difficult to record MMN changes to that many contrasts, as changes in MMN amplitude following HVPT have yet to be demonstrated for as many contrasts. However, this would be an interesting and worthwhile direction to explore.

5.4. Conclusion

The experiments reported in the current thesis have demonstrated that the MMN is an adequate tool for exploring perception of multiple phonetic contrasts. The length of testing multiple contrasts is reduced by using a paradigm design that allows for alternating the phonemes in the standard and the deviant stimuli. Additionally, the mTRF analysis framework generates sufficient statistical power even if fewer deviant instances are included. MMNs analysed with the mTRF framework show cross-language effects, similar to those reported using conventional MMN methodologies. Nevertheless, the current thesis demonstrates that neural responses to IPA vowels do not display a pattern of MMN elicitation resembling classic phonetic vowel maps for either English nor Spanish speakers. Perhaps, replacing them with native vowel exemplars would create the desired classic vowel maps for both speaker groups. In contrast, the memory-card HVPT investigation showed that the memory-card game helped Spanish-Catalan primary school children learn the southern British English vowel set. Moreover, it demonstrated a bilingual advantage in phonetic learning in the older years of primary school. Altogether, these methodologies offer an optimal way of observing phonetic learning across the lifespan and will allow researchers to investigate the development of phonetic abilities and the impact of bi- and multilingualism in native and non-native speech processing.

6. References

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America*, *101*(2), 1090-1105. https://doi.org/10.1121/1.418031

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of Bilinguals Over Monolinguals in Learning a Third Language. *Bilingual Research Journal*, *33*(2), 173-199. https://doi.org/10.1080/15235882.2010.502797

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

Amengual, M. (2016). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*, *20*(2), 133-152. https://doi.org/10.1177/1367006914544988

Amengual, M., & Chamorro, P. (2015). The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts. *Phonetica*, *72*(4), 207-236. https://doi.org/10.1159/000439406

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning*. *Bilingualism: Language and Cognition*, *18*(4), 683-695. https://doi.org/10.1017/S1366728914000777

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., &

Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Best, C. (1995). A direct realist view of crosslanguage speech perception. En *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204).

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345-360. https://doi.org/10.1037/0096-1523.14.3.345

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. En *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege* (pp. 13-34). https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2018). Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00081

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile | An Easy-to-Use Instrument to Assess Bilingualism*. https://sites.la.utexas.edu/bilingual/

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303-328. https://doi.org/10.1017/S0142716400008912

Borst, J. P., Schneider, D. W., Walsh, M. M., & Anderson, J. R. (2013). Stages of Processing in Associative Recognition: Evidence from Behavior, EEG, and Classification. *Journal of Cognitive Neuroscience*, *25*(12), 2151-2166. https://doi.org/10.1162/jocn_a_00457

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy*, *2*(1), 29-49. https://doi.org/10.1207/S15327078IN0201_3

Bosch, L., & Sebastian-Galles, 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), 217-244.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272-284. https://doi.org/10.1121/1.1487837

Brandmeyer, A., Sadakata, M., Spyrou, L., McQueen, J. M., & Desain, P. (2013). Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2013.00265

Brekelmans, G. (2020). *Phonetic vowel training for child second language learners: The role of input variability and training task*. 460.

Broderick, M. P., Di Liberto, G. M., Anderson, A. J., Rofes, A., & Lalor, E. C. (2021). Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84597-9

Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading & Writing*, *16*(6), 573-609.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(3). https://doi.org/10.1017/S0142716407070257

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science*, *21*(3), 343-348.

Bylund, E., Hyltenstam, K., & Abrahamsson, N. (2021). Age of acquisition – not bilingualism – is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism*, *24*(1), 18-30. https://doi.org/10.1017/S1366728920000188

Casillas, J. (2015). Production and Perception of the /i/-/I/ Vowel Contrast: The Case of L2-Dominant Early Learners of English. *Phonetica*, *72*(2-3), 182-205. https://doi.org/10.1159/000431101

Cebrian, J., & Carlet, A. (2014). Second-Language Learners' Identification of Target-Language Phonemes: A Short-Term Phonetic Training Study. *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, *70*(4), 474-499.

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), Article 5. https://doi.org/10.1038/1561

Chládková, K., Escudero, P., & Lipski, S. C. (2013). Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain and Language*, *126*(3), 243-252. https://doi.org/10.1016/j.bandl.2013.05.020

Collins, B. S., & Mees, I. M. (2013). *Practical Phonetics and Phonology: A Resource Book for Students*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/ucl/detail.action?docID=1125249

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2021.705621

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. The MIT Press. https://doi.org/10.7551/mitpress/9012.001.0001

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*(4), 568-581. https://doi.org/10.1016/j.wocn.2012.05.001

Davine, M., Tucker, G. R., & Lambert, W. E. (1971). The perception of phoneme sequences by monolingual and bilingual elementary school children. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, *3*(1), 72-76. https://doi.org/10.1037/h0082251

de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, *171*(2), 331-339. https://doi.org/10.1016/j.jneumeth.2008.03.015

de Jong, G., McDougall, K., & Nolan, F. (2007). Sound Change and Speaker Identity: An Acoustic Study. En *Speaker Classification II* (pp. 130-141). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74122-0_12

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*(4), 919-924. https://doi.org/10.1097/00001756-199703030-00021

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Enomoto, K. (1994). L2 Perceptual Acquisition: The Effect of Multilingual Linguistic Experience on the Perception of a «Less Novel» Contrast. *Edinburgh Working Papers in Applied Linguistics*. https://eric.ed.gov/?id=ED373553

Escudero, P., & Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, *26*(4), 551-585.

Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2017a). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, *20*(1), e12427. https://doi.org/10.1111/desc.12427

Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2017b). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, *20*(1), e12427. https://doi.org/10.1111/desc.12427

Flege, J. (1995). Second language speech learning: Theory, findings and problems. En *Speech perception and linguistic experience: Issue in cross-language research* (pp. 229-273). Timonium, MD: York Press.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589-1608. https://doi.org/10.1121/1.406818

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425-442. https://doi.org/10.1017/S0142716400066029

Flege, J. E. (2008). 9. Give Input a Chance! En *Input Matters in SLA* (pp. 175-190). Multilingual Matters. https://www-degruyterbrill-com.libproxy.ucl.ac.uk/document/doi/10.21832/9781847691118-012/html

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). En R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437-470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving Vowels in a Second Language. *Studies in Second Language Acquisition*, *26*(1), 1-34.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973-2987. https://doi.org/10.1121/1.428116

Flege, J. E., Mackay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567-598. https://doi.org/10.1017/S0142716402004046

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491. https://doi.org/10.1016/S0167-6393(02)00128-0

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*. SAGE Publications.

Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017). Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks. *The Journal of the Acoustical Society of America*, *142*(2), 1025-1033. https://doi.org/10.1121/1.4998706

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546-557. https://doi.org/10.1016/j.wocn.2011.07.002

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936-944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453-463. https://doi.org/10.1016/j.clinph.2008.11.029

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177-190. https://doi.org/10.1037/0012-1649.40.2.177

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209. https://doi.org/10.7717/peerj.3209

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013a). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201-218. https://doi.org/10.1177/0265659012467473

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013b). Phonetic and orthographic cues are weighted in speech sound perception by second language speakers: Evidence from Greek speakers of English. *The Journal of the Acoustical Society of America*, *134*(5), 4031-4031. https://doi.org/10.1121/1.4830717

Gonzalez-Ardeo, J. M. (2001). Engineering students and ESP in the Basque Country: SLA versus TLA. En *Looking beyond second language acquisition: Studies in tri-and multilingualism* (In J. Cenoz, B. Hufeisen, U. Jessner (eds.), pp. 75-95). Tubingen. https://go-gale-com.libproxy.ucl.ac.uk/ps/i.do?p=AONE&u=ucl_ttda&id=GALE|A106585392&v=2.1&it=r

Goriot, C., McQueen, J. M., Unsworth, S., Hout, R. van, & Broersma, M. (2020). Perception of English phonetic contrasts by Dutch children: How bilingual are early-English learners? *PLOS ONE*, *15*(3), e0229902. https://doi.org/10.1371/journal.pone.0229902

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, *12*(2), 91-114. https://doi.org/10.1016/S0095-4470(19)30858-7

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00279

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131-149. https://doi.org/10.1017/S136672899800025X

Grosjean, F. (2015). The Complementarity Principle and its impact on processing, acquisition, and dominance. En C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 66-84). Cambridge University Press. https://doi.org/10.1017/CBO9781107375345.004

Hansen Edwards, J. G., & Zampini, M. L. (2008). *Phonology and Second Language Acquisition*. John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/ucl/detail.action?docID=622973

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377-396. https://doi.org/10.1006/jpho.2000.0121

Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *The Journal of the Acoustical Society of America*, *124*(4), 2291-2302. https://doi.org/10.1121/1.2967472

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish /t-tː/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics*, *38*(4), 594-603. https://doi.org/10.1016/j.wocn.2010.08.005

Herschensohn, J. (Ed.). (2007). Pressed for time: Age constraints in second language acquisition. En *Language Development and Age* (pp. 134-171). Cambridge University Press. https://doi.org/10.1017/CBO9780511486487.006

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research*, *1360*, 89-105. https://doi.org/10.1016/j.brainres.2010.08.092

Højlund, A., Horn, N. T., Sørensen, S. D., McGregor, W. B., & Wallentin, M. (2022). Foreign language learning and the mismatch negativity (MMN): A longitudinal ERP study. *Neuroimage: Reports*, *2*(4), 100138. https://doi.org/10.1016/j.ynirp.2022.100138

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105-120. https://doi.org/10.1016/j.wocn.2015.06.007

Iverson, P., & Evans, B. G. (2009a). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866-877. https://doi.org/10.1121/1.3148196

Iverson, P., & Evans, B. G. (2009b). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakersa). *J. Acoust. Soc. Am.*, *126*(2).

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278. https://doi.org/10.1121/1.2062307

Iverson, P., Herrero, B. P., & Katashima, A. (2023). *Memory-card vowel training for child and adult second-language learners: A first report.*

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, *97*(1), 553-562. https://doi.org/10.1121/1.412280

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145-160. https://doi.org/10.1017/S0142716411000300

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J. W., & Desimone, R. (2004). Human Posterior Auditory Cortex Gates Novel Sounds to Consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809-6814.

Jensen, B. (2008). Metalinguistic Awareness. En *Encyclopedia of Bilingual Education* (Vol. 2, pp. 551-554).

Junttila, K., Smolander, A.-R., Karhila, R., Giannakopoulou, A., Uther, M., Kurimo, M., & Ylinen, S. (2022). Gaming enhances learning-induced plastic changes in the brain. *Brain and Language*, *230*, 105124. https://doi.org/10.1016/j.bandl.2022.105124

Kasisopa, B., El-Khoury Antonios, L., Jongman, A., Sereno, J. A., & Burnham, D. (2018). Training Children to Perceive Non-native Lexical Tones: Tone Language Background, Bilingualism, and Auditory-Visual Information. *Frontiers in Psychology*, *9*, 1508.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The Influence of Meaning on the Perception of Speech Sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11381-11386.

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports (Nature Publisher Group)*, *6*, 19741. https://doi.org/10.1038/srep19741

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, Rate and Eventual Attainment in Second Language Acquisition. *TESOL Quarterly*, *13*(4), 573-582. https://doi.org/10.2307/3586451

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central Auditory System Plasticity Associated with Speech Discrimination Training. *Journal of Cognitive Neuroscience*, *7*(1), 25-32. https://doi.org/10.1162/jocn.1995.7.1.25

Krøjgaard, P., Sonne, T., Lerebourg, M., Lambek, R., & Kingo, O. S. (2019). Eight-year-olds, but not six-year-olds, perform just as well as adults when playing Concentration: Resolving the enigma? *Consciousness and Cognition*, *69*, 81-94. https://doi.org/10.1016/j.concog.2019.01.015

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. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (1993). Innate Predispositions and the Effects of Experience in Speech Perception: The Native Language Magnet Theory. En B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 259-274). Springer Netherlands. 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), 831-844. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic Learning as a Pathway to Language: New Data and Native Language Magnet Theory Expanded (NLM-e). *Philosophical Transactions: Biological Sciences*, *363*(1493), 979-1000.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13-F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1-19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227-247. https://doi.org/10.1017/S0142716405050150

Lengeris, A., & Nicolaidis, K. (2015). Effect of phonetic training on the perception of English consonants by Greek speakers in quiet and noise. *Proceedings of Meetings on Acoustics*, *22*(1), 060002. https://doi.org/10.1121/2.0000025

Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Versión 1.10.7) [Software]. https://cran.r-project.org/web/packages/emmeans/index.html

Lim, S., & Holt, L. L. (2011). Learning Foreign Sounds in an Alien World: Videogame Training Improves Non-Native Speech Categorization. *Cognitive Science*, *35*(7), 1390-1405. https://doi.org/10.1111/j.1551-6709.2011.01192.x

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886. https://doi.org/10.1121/1.1894649

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-Related Potentials and the Categorical Perception of Speech Sounds. *Ear & Hearing*, *16*(1), 68-89.

Masapollo, M., Polka, L., & Ménard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition*, *166*, 358-370. https://doi.org/10.1016/j.cognition.2017.06.001

McCarthy, K. M., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech Perception and Production by Sequential Bilingual Children: A Longitudinal Study of Voice Onset Time Acquisition. *Child Development*, *85*(5), 1965-1980.

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*, *9*(1), Article 1. http://dx.doi.org/10.1038/s41598-019-55085-y

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, *38*(3), 581-605. https://doi.org/10.1177/0267658320988061

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

Muñoz, C. (2006). Chapter 1 .The Effects of Age on Foreign Language Learning: The BAF Project. En C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 1-40). Multilingual Matters. https://www.degruyterbrill.com/document/doi/10.21832/9781853598937-003/html

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), Article 6615. https://doi.org/10.1038/385432a0

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007a). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007b). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140-144. https://doi.org/10.1016/j.clinph.2003.04.001

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. https://doi.org/10.1037/0096-1523.24.3.756

Nittrouer, S. (1996). The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM. *Journal of Speech and Hearing Research*, *39*(5), 1059-1070. https://doi.org/10.1044/jshr.3905.1059

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, *101*(4), 2253-2266. https://doi.org/10.1121/1.418207

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *The Journal of the Acoustical Society of America*, *102*(6), 3711-3722. https://doi.org/10.1121/1.420135

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*. Gale Academic OneFile. https://link.gale.com/apps/doc/A284222352/AONE?u=ucl_ttda&sid=bookmark-AONE&xid=6de44ac2

Partanen, E., Pakarinen, S., Kujala, T., & Huotilainen, M. (2013). Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clinical Neurophysiology*, *124*(8), 1578-1585. https://doi.org/10.1016/j.clinph.2013.02.014

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25-28. https://doi.org/10.1016/j.neulet.2003.08.013

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285-290. https://doi.org/10.3758/BF03213946

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*(1), 37-52. https://doi.org/10.3758/BF03206758

Rabinovitch, M. S., & Parver, L. M. (1966). *Auditory discrimination in monolinguals and poliglots*. meeting of the Canadian Psychological Association, Montreal, Canada.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*(2), 162-172. https://doi.org/10.1111/j.1467-7687.2005.00403.x

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and Multimodal High Variability Phonetic Training: Potential, Limits, and Future Directions. *Language Learning*, *72*(4), 1049-1091. https://doi.org/10.1111/lang.12503

Saito, K., Kachlicka, M., Suzukida, Y., Petrova, K., Lee, B. J., & Tierney, A. (2022). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning. *Cognition*, *229*, 105236. https://doi.org/10.1016/j.cognition.2022.105236

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224. https://doi.org/10.1017/S0142716417000418

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-Related Potentials to Infrequent Changes in Synthesized Phonetic Stimuli. *Journal of Cognitive Neuroscience*, *2*(4), 344-357. https://doi.org/10.1162/jocn.1990.2.4.344

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.

Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., Castillo, D., & Strange, W. (2021). The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-language learners of English. *Bilingualism*, *24*(4), 642-655. https://doi.org/10.1017/S1366728921000201

Sharma, A., & Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *The Journal of the Acoustical Society of America*, *104*(1), 511-517. https://doi.org/10.1121/1.423252

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078-1083. https://doi.org/10.1121/1.428048

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, *107*(5), 2697-2703. https://doi.org/10.1121/1.428655

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(1), 64-71. https://doi.org/10.1016/0168-5597(93)90029-O

Shinohara, Y., & Iverson, P. (2013). Computer-based English /r/-/l/ perceptual training for Japanese children. *The Journal of the Acoustical Society of America*, *133*(5), 3333-3333. https://doi.org/10.1121/1.4805599

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, *66*, 242-251.

Shinohara, Y., & Iverson, P. (2021a). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*, 101108-.

Shinohara, Y., & Iverson, P. (2021b). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*. https://doi.org/10.1016/j.wocn.2021.101108

Shtyrov, Y., & Pulvermüller, F. (2007). Language in the Mismatch Negativity Design: Motivations, Benefits, and Prospects. *Journal of Psychophysiology*, *21*(3-4), 176-187. https://doi.org/10.1027/0269-8803.21.34.176

Simonet, M., & Amengual, M. (2020). Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism*, *24*(2), 208-221. https://doi.org/10.1177/1367006919826388

Snowling, M., Hulme, C., Butterworth, B., Morton, J., Snowling, M. J., & Warrington, E. K. (1997). The development of phonological skills. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *346*(1315), 21-27. https://doi.org/10.1098/rstb.1994.0124

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273-294. https://doi.org/10.1177/0023830909357156

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*, 163-170. https://doi.org/10.1016/j.cognition.2018.06.001

Spinu, L. E., Hwang, J., & Lohmann, R. (2018). Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English. *International Journal of Bilingualism*, *22*(3), 350-370. https://doi.org/10.1177/1367006916681080

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456-466. https://doi.org/10.1016/j.wocn.2010.09.001

Sundara, M., & Polka, L. (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition*, *106*(1), 234-258. https://doi.org/10.1016/j.cognition.2007.01.011

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*(1), 232-242. https://doi.org/10.1016/j.cognition.2007.12.013

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505-513. https://doi.org/10.1016/j.wocn.2010.08.006

Sussman, J. E., & Carney, A. E. (1989). Effects of Transition Length on the Perception of Stop Consonants by Children and Adults. *Journal of Speech, Language, and Hearing Research*, *32*(1), 151-160. https://doi.org/10.1044/jshr.3201.151

Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, *62*(4), 1231-1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport*, *9*(16), 3557-3560. https://doi.org/10.1097/00001756-199811160-00003

Tremblay, M.-C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *The Journal of the Acoustical Society of America*, *132*(5), 3465-3474. https://doi.org/10.1121/1.4756955

Tsogli, V., Sebastian, J., Tatsuya, D., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports (Nature Publisher Group)*, *9*(1). https://doi.org/10.1038/s41598-019-42066-4

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179-191. https://doi.org/10.1002/dev.21179

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263-290. https://doi.org/10.1016/j.wocn.2004.10.002

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665-3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, *118*(4), 2618-2633. https://doi.org/10.1121/1.2011156

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043. https://doi.org/10.1121/1.1531176

Wang, Y., & Kuhl, P. K. (2003). *Evaluating the ``Critical Period" Hypothesis: Perceptual Learning of Mandarin Tones in American Adults and American Children at 6, 10 and 14 Years of Age*. 4.

Werker, J. F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics*, *7*(2), 141-155. https://doi.org/10.1017/S0142716400007360

Werker, J. F. (1989). Becoming a Native Listener. *American Scientist*, *77*(1), 54-59.

Werker, J. F., & and Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197-234. https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35-44. https://doi.org/10.3758/BF03207136

Werker, J. F., & Tees, R. C. (1983a). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *37*(2), 278-286. https://doi.org/10.1037/h0080725

Werker, J. F., & Tees, R. C. (1983b). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *37*(2), 278-286. https://doi.org/10.1037/h0080725

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866-1878. https://doi.org/10.1121/1.390988

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3), 233-251. https://doi.org/10.1002/dev.20060

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3-4), 147-163.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638-642. https://doi.org/10.1111/1469-8986.3650638

Wottawa, J., Adda-Decker, M., & Isel, F. (2022). Neurophysiology of non-native sound discrimination: Evidence from German vowels and consonants in successive French–German bilinguals using an MMN oddball paradigm. *Bilingualism*, *25*(1), 137-147. https://doi.org/10.1017/S1366728921000468

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience*, *22*(6), 1319-1332.

Zhang, J., Dennis, J. M., & Houseman, C. (2018). The role of family and community bicultural socialization in the bilingual proficiency of immigrant young adults. *International Journal of Intercultural Relations*, *67*, 44-57. https://doi.org/10.1016/j.ijintrel.2018.09.003

Zhang, L., Li, Z., Zhang, F., Gu, R., Peng, W., & Hu, L. (2020). Demystifying signal processing techniques to extract task-related EEG responses for psychologists. *Brain Science Advances*, *6*(3), 171-188. https://doi.org/10.26599/BSA.2020.9050018

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, *46*(1), 226-240. https://doi.org/10.1016/j.neuroimage.2009.01.028

Zinszer, B. D., Yuan, Q., Zhang, Z., Chandrasekaran, B., & Guo, T. (2022). Continuous speech tracking in bilinguals reflects adaptation to both language and noise. *Brain and Language*, *230*, 105128. https://doi.org/10.1016/j.bandl.2022.105128