UNIVERSITY OF CALIFORNIA, MERCED

How top-down and bottom-up factors modulate bilingual phonemic encoding

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Cognitive and Information Sciences

by

Zunaira J. Iqbal

Committee in charge:

Kristina C. Backer, Ph.D., Chair

Michael J. Spivey, Ph.D.

Antoine J. Shahin, Ph.D.

Heather Bortfeld, Ph.D.

2024

The Dissertation of Zunaira Jahan Iqbal is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Michael J. Spivey, Ph.D.

Cognitive & Information Sciences, UC Merced

Antoine J. Shahin, Ph.D.

Cognitive & Information Sciences, UC Merced

Heather Bortfeld, Ph.D.

Psychological Sciences, UC Merced

Kristina C. Backer, Ph.D., Chair

Cognitive & Information Sciences, UC Merced

University of California, Merced

2024

This dissertation is dedicated to Mama, Baba, Aaron, Usman, Umar, and Ammi.

I am forever grateful for your endless support and love.

Table of Contents

List Figures & Tables

List of Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 1

Chapter 2

Chapter 3

Acknowledgements

Funding

National Institute on Deafness and Other Communications Disorders Grant R01 DC-013543

National Science Foundation National Research Traineeships in Intelligent Adaptive Systems

Georgiana Ducas Endowment Fund Scholar (ARCS)

Graduate Dean's Dissertation Fellowship

Collaborators/Co-Authors

Alejandra E. Santoyo, Cognitive & Information Sciences, UC Merced

Mariel G. Gonzales, Cognitive & Information Sciences, UC Merced

Ramesh Balasubramaniam, Cognitive & Information Sciences, UC Merced

Personal Acknowledgements

To my labmates, Alex Santoyo, Mariel Gonzales, Adam Holm, Cerena Lee, Madison Lacanlale, and Achint Sharma—thank you all for being the best distraction when coming to campus. You have all been the best listeners, the greatest supporters, and the kindest friends. Especially to the original trio, Alex and Mariel, I wouldn't have been able to make it through graduate school without you two. Thanks for the endless memories and laughs over the last five years!

To my committee, Kristina Backer, Antoine Shahin, Heather Bortfeld, and Michael Spivey—thanks for all the invaluable guidance and patience. Especially to Kristina and Tony for putting up with years of tears, phone calls, and meetings.

To all the faculty and graduate students of UCM CogSci, especially Benny Nguyen and Helia Hosseinpour—thank you for all the fun times and support!

To my dear friends, Puja, Sania, Humayra, and Yasmin—you girls are the greatest friends. Thank you for always supporting me, listening to my rants, and keeping our friendship strong!

To my parents, Sarwat and Asif, for always supporting me. I couldn't have gotten here without your endless help and love. To my brothers, Usman and Umar, for always hanging out with me and making me laugh over the dumbest things. Thank you to my grandma, Ammi, who has been my biggest supporter and always reminds me of how proud she is of my accomplishments.

Lastly, thank you to my husband, Aaron, who has always believed in me and supported me from day one.

Zunaira J. Iqbal

Email: ziqbal2@ucmerced.edu

Website: www.zunairaiqbal.com

Interests

My research interest regards spoken language perception, especially as it relates to bilingual speakers. Specifically, I examine how different contexts, such as visual, attentional, and lexico-semantic cues, impact bilinguals' especially phonemic encoding. I use both behavioral and EEG measures to carry out my research.

Education

PhD candidate in Cognitive and Information Sciences Aug 2019-Jul 2024

University of California, Merced

- Advisors: Kristina Backer, Michael Spivey, & Antoine Shahin
- Became candidate in Spring 2023
- Expected Graduation Date: Summer 2024

Bachelor of Science, Psychology-Biology Emphasis June 2019

University of California, Davis

- Minors in Linguistics and Spanish
- Graduated with High Honors

Peer reviewed publications

Santoyo, A.*, Gonzales, M. G.*, **Iqbal, Z. J.***, Backer, K. C., Balasubramaniam, R., Bortfeld, H., & Shahin, A. J. (2023). *Neurophysiological time course of timbre-induced melodic perception.* JNeurophysiol 1;130(2):291-302. https://doi.org/10.1152/jn.00042.2023 (Selected by the American Physiological Society for a distinction in scholarship)
*Equal first authors

Iqbal, Z. J., Shahin, A. J., Bortfeld, H., & Backer, K. C. (2023). The McGurk Illusion: A Default Mechanism of the Auditory System. *Brain Sciences*, *13*(3), 510. https://doi.org/10.3390/brainsci13030510

Conference Abstracts

Iqbal, Z. J., Shahin, A. J., Spivey, M. J., & Backer, K. C. (2023, November 16-19). *Neural Mechanisms of Phonemic Restoration in Bilingual Listeners* [Poster presentation]. Psychonomic Society 64th Annual Meeting, San Francisco, CA.

Iqbal, Z. J., Backer, K. C., Lacanlale, M. C., Bortfeld, Heather, & Shahin, A. J. (2022, November 17-20). *Does the McGurk Illusion Arise from Ambiguity?* [Poster presentation]. Psychonomic Society 63rd Annual Meeting, Boston, MA, USA.

Santoyo, A., Gonzales, M. G., **Iqbal, Z. J.**, Backer, K. C., Balasubramaniam, R., Bortfeld, H., & Shahin, A. J. (2022, November 12-16). *Neural encoding of melody to pitch less sounds* [Poster presentation]. Society for Neuroscience 50th Annual Meeting, San Diego, CA, USA.

Iqbal, Z. J., Shahin, A. J., & Backer, K. C. (2022, October 6-8). *Perception of Audiovisual Pairings Among Heritage Speakers of Spanish* [Poster presentation]. The Fourteenth Annual Society for the Neurobiology of Language Meeting, Philadelphia, PA, USA.

Iqbal, Z. J., Shahin, A. J., Spivey, M. J., & Backer, K. C. (2021, November 4-7). The Influence of Visual Information on Phonetic Representations in Spanish-English Bilinguals [Poster presentation]. Psychonomic Society 62nd Annual Meeting, Virtual Conference.

Iqbal, Z. J., Shahin, A. J., & Backer, K. C. (2021, October 5-8). Spanish-English Bilinguals' Phonetic Perception in Speech-Shaped Noise [Poster presentation]. 13th Annual Meeting of the Society for the Neurobiology of Language, Virtual Conference.

Iqbal, Z. J. (2021). How Speech-Shaped Noise Influences Phonetic Perception in Monolingual and Bilingual Listeners. UC Davis 7th Annual Symposium on Language Research. Davis, CA.

Iqbal, Z. J. & Corina, D.P. (2019). The Syntax-Space Effect in American Sign Language. UC Davis 30th Talk presented at the Annual Undergraduate Research, Scholarship & Creative Activities Conference. Davis, CA.

Iqbal, Z. J. (2018). *Examining L2 Urdu in the United States.* Talk presented at the UC Davis ME/SA Research Symposium. Davis, CA.

Invited Talks

The McGurk Illusion: A Default Mechanism of the Auditory System, University of California, Irvine, Center for Hearing Research, May 24, 2023

Teaching Experience

Teaching Assistant

COGS 005: Introduction to Language and Linguistics	Spring 2020, 2022-2023
COGS 142: Audition	Fall 2019, Fall 2021, Fall 2023

Externally Funded Grants

Graduate Dean's Dissertation Fellowship, *$10,875*	Summer 2024
ARCS Scholar, *$13,300*	Fall 2023-Spring 2024
ARCS Scholar, *$13,300*	Summer 2023
NRT Intelligent Adaptative Systems Fellowship, *$26,600*	Spring 2022-Summer 2022
NRT Intelligent Adaptative Systems Fellowship, *$34,000*:	Fall 2020-Spring 2021

Awards

Publication *Neurophysiological time course of timbre-induced melodic perception* (2023) selected by the American Physiological Society for a distinction in scholarship.

Service

External Vice-President of UC Merced's Graduate Student Association	Fall 2020-Spring 2021
Organizer of Department's Women's Tea	Fall 2020-Current
Mentor for W-STEM	Fall 2020
Social Chair of Department's Graduate Student Group	Fall 2020
Linguistics Society of America Summer Institute Volunteer	Summer 2019

Methodological Skills

EEG, MATLAB, R, Presentation (Neurobehavioral Systems), audiometry testing

Membership

Society for Neuroscience, Member, 2022-Present

Psychonomic Society, Member, 2020-Present

Society for the Neurobiology of Language, Member, 2020-Present

Languages

English, Urdu/Hindi

Abstract

With the rise of bilinguals globally, extensive research has been conducted to understand how bilingualism affects cognitive functions differently from monolingualism, with significant implications for bilingual education, healthcare, and other fields. Specific to language processing, differences have been identified between second language (L2) speakers and monolinguals, particularly in contexts such as speech perception in noise and executive control. Despite the advances in this field of language processing, there remains a critical gap in understanding how bilinguals' phonological awareness of their first language (L1) and L2 influences their phonetic encoding and categorization. In this dissertation, I examine how various top-down and bottom-up factors modulate phonemic encoding in bilinguals, both behaviorally and neurophysiologically. The first chapter investigates how ambiguous audiovisual situations result in a default phonetic perception, providing a foundation for the second chapter. The second chapter explores how visual information influences the phonetic encoding of specific phonemes by Spanish-English bilinguals. Lastly, the third chapter addresses the influence of selective attention, lexicosemantic encoding, and low-level sound encoding on bilingual phonetic encoding, focusing on how bilinguals respond to phonemic restoration. This experiment is conducted using electroencephalography (EEG) to provide insights into the neurophysiological mechanisms underlying these processes. By integrating behavioral and neurophysiological approaches, this dissertation aims to uncover the complex mechanisms underlying phonetic encoding in bilinguals, contributing to a comprehensive understanding of bilingual cognitive processing and its broader implications.

Introduction

More than half of the world's population can be considered bilingual. A bilingual is an individual that is able to speak two or more languages or dialects in their everyday life (Ansaldo et al., 2008; Grosjean, 1994). In the US alone, there are over 67 million bilingual individuals, and this number only continues to grow (U.S. Census Bureau, 2018). In past decades, a substantial amount of research has been conducted to understand the consequences of being bilingual and how bilinguals process various cognitive functions differently from monolinguals (Bialystok, 2010; Bialystok et al., 2012; Kroll et al., 2014). This comparison is significant as it influences approaches taken towards bilingual education, healthcare, and more. Specific to language processing, research has identified differences between second language (L2) speakers and monolinguals of the same language, across various contexts (e.g., speech in noise perception, executive control, etc.) (Bialystok et al., 2012; Kroll & Bialystok, 2013; Skoe & Karayanidi, 2019). Despite the knowledge acquired so far, there is still limited data on how bilinguals' phonological awareness of their first language (L1) and L2 influences their phonetic encoding and categorization. Phonological awareness refers to the set of phonemes and phonological rules specific to a language.

This is a critical gap within the literature that needs to be researched because each individual language has its own set of phonemes and phonological rules, that may or may not overlap with that of other languages. A clear example of this linguistic conflict is in the case of Spanish phonology compared to that of English; certain phonemes that are differentiated in English are mapped onto a single Spanish phoneme. For example, the phonemes /v/ and /b/ are perceived as two distinct sounds in English, but in Spanish /v/ is mapped onto /b/ (see Figure 1). Consequently, many native Spanish speakers replace /v/ sounds in English words with /b/ (e.g., *volleyball* is often pronounced *bolleyball*). Another example distinguishing Spanish and English is the case of /s/ and /z/. Again, in English these are perceived as two separate phonemes. However, in Spanish, /z/ gets mapped onto /s/. This difference between the two languages raises the question of how Spanish-English bilinguals perceptually and neurophysiologically represent these phonemes, shedding light on the effects of bilingualism on language perception. Of particular interest are heritage Spanish speakers in the United States: Spanish-English bilinguals who learned Spanish in the home as their first language and learned English outside the home.

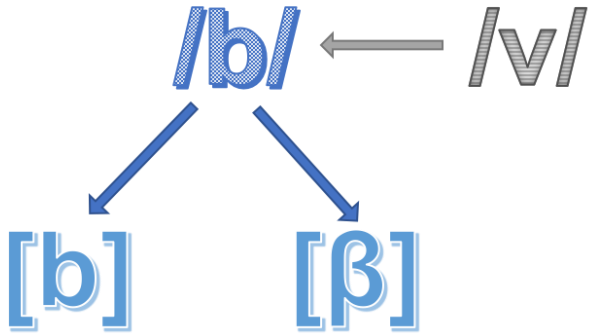

Figure 1: A depiction of how in Spanish phonology, the phoneme /v/ is mapped onto /b/. Spanish has two allophones of /b/: [b] and [β]. It is important to note that Spanish does orthographically use 'v', but the symbol represents the phoneme /b/.

The goal of my dissertation is to dissect bilingual phonetic encoding, behaviorally and neurophysiologically, with a focus on how aspects such as visual speech information, noise, and higher-level factors (i.e., attention, lexico-semantics) influence bilinguals' phonemic encoding. The present research focuses on Spanish-English bilinguals due to 1) the key differences between Spanish and English phonology mentioned above and, 2) the prevalent Hispanic population in California. My dissertation addresses these gaps through a model that aims to understand 1) the role of visual information on phonemic encoding, 2) how top-down information differs in phonemic encoding between bilinguals and monolinguals, and 3) how low-level acoustic features interfere in phonemic encoding.

Background

Phonetic Encoding and Phonemic Perception

It is important to clarify some key terminology before discussing the details of phonetic encoding and phonemic perception. At its core, the acoustic speech input received must be encoded into smaller units of language, or phonemes, before any sense of higher-level language comprehension can occur. This process is often referred to as phonetic encoding. Making sense of encoded phonemes, e.g., how they are heard, is referred to as phonemic perception (Kleinschmidt & Jaeger, 2015; Samuel, 1982). For the sake of this dissertation however, I will use the terms phonetic encoding and phonemic perception/encoding interchangeably, since the two processes, at the cortical level, are largely one and the same.

To achieve successful speech perception, listeners must accurately and correctly map the acoustic input onto the correct phonetic category (often also referred to as *phonetic prototype*) (Samuel, 1982; Samuel & Newport, 1979). How exactly these phonetic prototypes are established has been extensively debated. It is well known that infants before the age of 6-8 months are considered to be the universal listener (Kuhl et al., 1992; Werker et al., 1981); that is, they are able to discriminate between any phoneme in any language, regardless of the language(s) spoken by their primary caretaker(s). However, human infants around 8 months and older are suddenly unable to discriminate phonemes

outside their first language; their phonetic categories have already been established based off the language(s) spoken at home (Werker et al., 1981; Werker & Tees, 1984).

But what about the case of bilinguals, who face the challenge of the use of different speech sounds across the two languages they know? A phonetic cue that is informative for phonemic discrimination in one language may not be necessarily useful in another language and may map differentially onto the categorical organization of the monolingual/bilingual phonetic code (Abramson & Lisker, 1973; Polka et al., 2001). Depending on when a bilingual has learned their second language, their phonemic representations will vary (Sundara & Polka, 2008). Research suggests that bilinguals may have a double phonemic representation: two phonemic representations, one for each language, for one acoustic-event (Garcia-Sierra et al., 2009). To test this double phonemic effect, many studies have been conducted and have produced mixed results. Some experimental situations have provided evidence for double phonemic representations within bilinguals, whereas others have failed to replicate this finding; taken together, these results suggest that bilinguals can be sensitive to experiencing dual phonemic representations in certain language contexts (Caramazza et al., 1973; Elman et al., 1977; Flege & Eefting, 1987; Hazan & Boulakia, 1993; Williams, 1977).

A possible explanation of speech perception views the process from the perspective of motor control. The motor theory of speech perception postulates that speech perception is inherently linked to the production mechanisms involved in speech (Liberman et al., 1967; Liberman & Mattingly, 1985). The theory suggests that when speech sounds are heard, we mentally reproduce the corresponding motor movements involved in producing those sounds. By mentally moving the articulators that correspond to a speech sound, these gestures can be mapped onto features of that specific phoneme. Research suggests that simply listening to verbal stimuli results in activation of speech-related motor areas of the brain (Cheung et al., 2016; D'Ausilio et al., 2009; Galantucci et al., 2006; Kerzel & Bekkering, 2000). In regard to L2 speech perception, supporters of the motor theory of speech perception suggest that the use of two or more languages requires more neurocognitive demand placed on the individual (Abutalebi & Green, 2016). Due to these cognitive demands, it is expected that activation of motor areas will occur. When listening to non-native speech sounds, lip corticobulbar activation increases as the perceived familiarity of the speech sound decreases (Schmitz et al., 2019). In bilinguals, it has been found that there is increased motor activation in L2 speech perception that differs from L1 perception (Barragan et al., 2022; Schmitz et al., 2018). These findings suggest that motor theory might provide a framework for understanding the unique speech perception processes in bilinguals, particularly in how they manage increased neurocognitive demands. This relationship between motor and cognitive processes might be crucial for bilinguals, who navigate two languages with potentially conflicting phonetic categories.

Building off the bilingual phonetic categorization literature thus far, the question that is yet to be answered is how heritage bilinguals mentally represent and perceive phonemes that exist in one language but are mapped differently in another language. For example, native Japanese speakers have difficulty perceptually distinguishing between the phonemes /r/ and /l/, whereas a native English speaker is aware that these are two

distinct phonemes in their native language. While research has been done looking at Japanese-English bilinguals and their categorization of /r/ and /l/, this work has mostly used bilingual populations who acquired English as a second language (Ingvalson et al., 2012; MacKain et al., 1981; Miyawaki et al., 1975; Mochizuki, 1981). More research on heritage speakers' phonetic categorization needs to be done, as they have a unique circumstance of being fairly fluent in a "home" language and another language.

Phonetic Encoding in the Brain

It is well known in the neurolinguistic literature that the temporal lobe, specifically the superior temporal gyrus (STG) and superior temporal sulcus (STS), is the main brain region recruited in higher-order auditory processing (Binder & Price, 2000; Gernsbacher & Kaschak, 2003; Hickok & Poeppel, 2004; Scott et al., 2000). Specifically, the left STS responds to the presence of phonetic information, further recruiting the anterior STS only when speech is intelligible (Scott et al., 2000). Work by Chang et al. (2010) used electrocorticography (ECoG) to look at categorical speech perception in posterior STG (pSTG). Using a continuum of /ba/ to /da/ to /ga/, they found that the pSTG is organized and sensitive to phonetic categories, but not to gradual acoustic variation along the continuum. This finding suggests that pSTG is a critical region for higher-level phonological processing and phonemic discrimination. There is also evidence that phonemic perception is also reflected within Heschl's gyrus (HG). Abbott and Shahin (2018a) and Shahin et al, (2018b) showed that phonetic encoding/perception is indexed by the N1 auditory evoked potential (AEP), known to be generated from the cortex within and surrounding HG.

In conclusion, prior research has demonstrated the critical role of the STG, STS, and HG in phonetic processing and speech perception. These regions carry out specialized functions in distinguishing phonetic categories and encoding phonemic information, affirming their importance in higher-order auditory and linguistic processing.

Audiovisual Speech Perception

Speech perception is strongly multimodal, consisting of both the auditory and visual domains (discussed above), and less often also involves the motor modality (Wilson et al., 2004). When we are trying to understand what a talker is saying to us, not only are we listening to their words, but we are also seeing their mouth movements as they talk, integrating visual context with auditory input. Perhaps the most profound effect that emphasizes the role of visual information on language (i.e., phonetic) perception is the McGurk effect (McGurk & Macdonald, 1976). When participants are presented with an auditory stimulus of /ba/ paired with an incongruent video showing the mouth movements of the speaker producing /ga/, many report perceiving a fused sound such as "da." This robust illusion goes to show that while speech perception relies upon what we hear, what we hear is also influenced by what we see.

In noisy environments, we rely upon visual information to aid in clarifying incoming speech signals (Sumby & Pollack, 1954). The presence of visual lip movements

increases speech comprehension performance compared to auditory-only conditions, particularly when listening to speech in noise (Sumby & Pollack, 1954). When presented with degraded target speech, visual lip-reading can help fill in the gaps (Grant & Seitz, 2000). Bilinguals are known to have difficulty comprehending speech in their second language in adverse listening conditions (Borghini & Hazan, 2018; M. L. G. Lecumberri et al., 2010; Mayo et al., 1997).While bilinguals can understand their second language (L2) speech just as well as native speakers of that language in quiet conditions (Florentine, 1985; Mayo et al., 1997; Takata & Nábělek, 1990), they perform less accurately at speech recognition when listening to their L2 in the presence of background noise (Mayo et al., 1997).

Due to less familiarity with their L2 and having a more difficult time discerning their L2 in adverse conditions, bilinguals have been shown to rely more upon visual cues than their monolingual counterparts (Chen & Hazan, 2007; Marian et al., 2018; Sekiyama, 1994). In fact, infants raised in a bilingual environment are more likely to naturally gaze at the mouth of speakers than monolingual infants (Pons et al., 2015). Additionally, it has been found that bilinguals are much more likely to experience the McGurk illusion in their L2 than monolinguals, again due to their tendency to gaze at the mouth more (Marian et al., 2018). In quiet conditions, the presence of visual mouth movements can greatly enhance a bilingual's perceptual ability to distinguish between non-native phonemic contrasts in their L2 (Navarra & Soto-Faraco, 2007). Similarly, in noisy contexts, bilinguals greatly benefit from visual speech cues paired with speech in their L2 (Chauvin et al., 2024).

AV Speech Perception in the Brain

The brain regions often associated with audiovisual speech integration include the posterior superior temporal gyrus (pSTG) and sulcus (pSTS) (Beauchamp et al., 2004, 2010, p. 20; Erickson et al., 2014), middle STS (B. T. Miller & D'Esposito, 2005; Venezia et al., 2017), middle temporal gyrus, and superior parietal lobule (Molholm et al., 2006). These regions are commonly linked with AV fusion, as they are within networks associated with phonological processing (Binder & Price, 2000; Hickok & Poeppel, 2004; Mesgarani et al., 2014). It is well known that these regions receive both auditory and visual input, making these regions possible hubs for AV integration/fusion (Venezia et al., 2017; Zhu & Beauchamp, 2017).

However, evidence also suggests that AV integration occurs within sensory networks, via direct or indirect pathways. For example, in the case of indirect stimulation, it is hypothesized that the STS/G assumes a top-down supportive role, such that following the evaluation of a mismatch pair of AV percepts (Hocking & Price, 2008), the STS/G relays the outcome to low-level speech areas (Arnal et al., 2009; Blank & von Kriegstein, 2013; Venezia et al., 2017) to regulate encoding of phonemes from the various sensory inputs. Alternatively, a direct account would posit that differential activity observed within the STS/G as well as HG may index visually mediated changes in phonetic encoding (Besle et al., 2008; Shahin et al., 2018). That is, visual input modulates (activates and inhibits) different populations of neurons within these networks

to precisely encode the visually conveyed phonemes, leading to differences in observed activity.

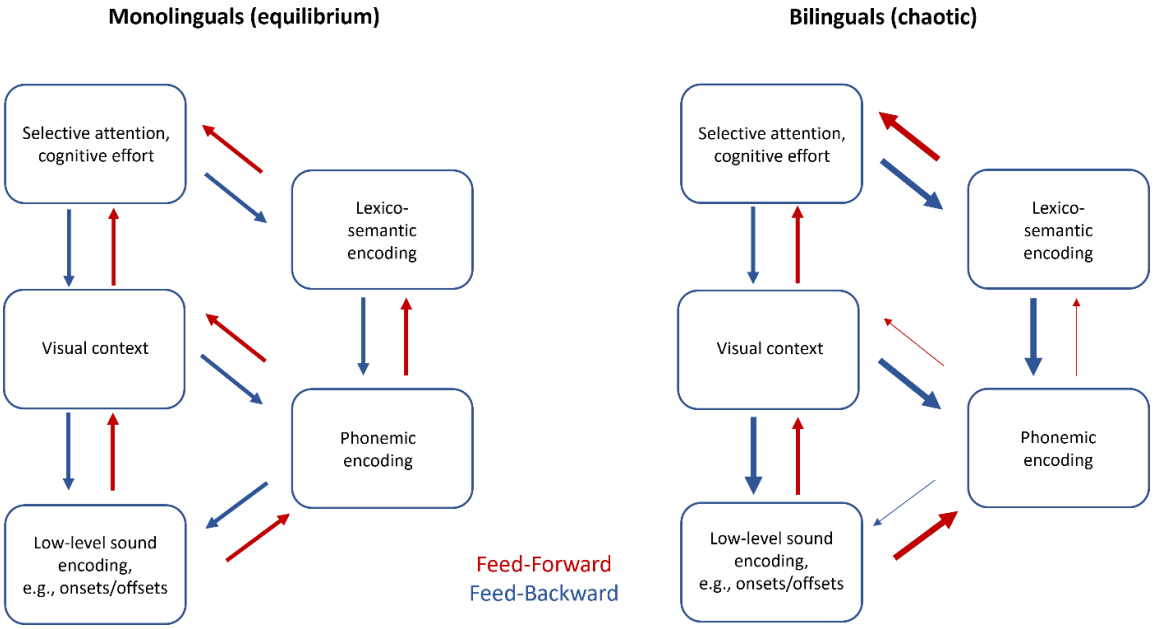

Figure 2: Theoretical framework of the feed-forward (bottom-up) and feed-backward (top-down) dynamics during spoken language processing in monolingual and bilingual populations.

Proposed Model

While research has been conducted on bilinguals and their phonemic perception, a gap still remains: how do bilinguals, who possess knowledge of two separate phonological systems, perceive and categorize certain phonemes, especially in different contexts? I propose a theoretical model, illustrated in Figure 2, to encapsulate how bilingual phonetic encoding/phonemic perception occurs, and how various factors influence this process.

In this proposed theoretical framework, bottom-up and top-down networks interact with each other in a feed-forward and feed-backward manner to perceptually organize phonetic, lexical, and semantic cues, optimizing spoken language comprehension. Bottom-up factors include acoustic characteristics, such as onsets and offsets of auditory input, and top-down factors include lexico-semantic knowledge, selective attention, and visual information. In the case of monolinguals, these various factors have equal weightings throughout the model. That is to say, monolinguals have an equilibrium of both feed-forward and -backward mechanisms. This equilibrium suggests adequate automatic processing with minimal cognitive, i.e., attentional effort.

However, bilinguals exhibit differential weightings with an imbalance across feed-forward and feed-backwards inputs. Starting with low-level encoding, bilinguals are less likely to filter extraneous information (such as acoustic onsets/offsets, noise, etc.), as implied by bilinguals' poor performance of speech perception in noisy situations (Mayo et al., 1997). This suggests then, that much more information, including unnecessary acoustic input, is sent upward, overwhelming or disrupting phonemic encoding. Moreover, since bilinguals have knowledge of the phonology of two languages, their phonemic encoding is much more cumbersome than that of a monolingual. That is, the process of encoding the correct phoneme is more difficult due to there being a larger phonetic inventory and possible overlaps across certain phonemes. This leads to weaker feed-forward input onto lexico-semantic templates. Due to weaknesses at the phonemic encoding and lexico-semantic matching stages, there is a heavier reliance upon information from selective attention. As for visual context, in situations whereby visual cues are present, the visual modality becomes a useful strategy in phonemic encoding facilitation. Visual context is known to greatly aid bilinguals' speech perception, possibly because of bilinguals' weak phonemic encoding (M. L. G. Lecumberri et al., 2010). In this framework, visual context works by perceptually organizing low level speech cues (i.e., acoustic characteristics of the input) by enhancing phonemic encoding, while also reducing reliance upon acoustic cues that are irrelevant to phonemic encoding.

Approach and Experiments

This dissertation comprises three experiments, to assess both the behavioral and neurophysiological processes underlying bilingual phonetic encoding.

Experiment 1: Audiovisual Ambiguity & the Mechanism Behind McGurk	How do ambiguous audiovisual situations result in defaulted phonetic perceptions? How does this explain the phonetic perception that arises from the McGurk Illusion?
Experiment 2: Audiovisual Speech Perception and Phonetic Encoding in Bilinguals	How do bilinguals represent and perceive phonemes that are represented differently in the two languages they speak? How does visual information influence this perception?
Experiment 3: Bilinguals and Phonemic Restoration	How do bilinguals experience the phonemic restoration effect? How does bottom-up (e.g., acoustic onsets/offsets) and top-down (i.e., attention) aid in their perception?

The goal of Experiment 1 is to understand the mechanism behind the McGurk illusion, and how ambiguous situations result in a "default" percept. While this study does not have a focus on bilinguals, it provides insight into the process behind

audiovisual speech perception, which is critical for understanding the distinction between bilinguals and monolinguals. Also, this experiment sets the stage for Experiment 2, which contrasts processing between monolinguals and bilinguals for ambiguous McGurk-like stimuli. This study fits into the proposed framework by delving into the interplay between visual and auditory representations on speech perception during ambiguous situations. As in the case of McGurk (auditory-/ba/ paired with visual-/ga/, perceiving /da/ta/tha/), when a viseme is ambiguous and indiscernible, this results in a failure of phonemic encoding on behalf of the visual cortex (i.e., visual context in the model). Because of this, the auditory system is left to default to naturally dominant and highly weighted percept.

The aim of Experiment 2 is to understand how visual information influences the phonetic encoding of certain phonemes by Spanish-English bilinguals. As discussed earlier, the presence of visual information can greatly influence our speech perception, especially in adverse and ambiguous listening situations (McGurk & Macdonald, 1976; Sumby & Pollack, 1954). While bilinguals can understand L2 speech just as well as native speakers of that language in quiet conditions (Florentine, 1985; Mayo et al., 1997; Takata & Nábělek, 1990), they perform less accurately at speech recognition when listening to their L2 in the presence of background noise (Mayo et al., 1997). Due to less familiarity in their L2 and having a more difficult time discerning their L2 in adverse conditions, bilinguals rely more heavily upon visual cues than their monolingual counterparts (Chen & Hazan, 2007; Marian et al., 2018; Navarra & Soto-Faraco, 2007; Sekiyama, 1994). Knowing the influence of visual information in bilinguals and how it can enhance L2 perception, how does the presence of mouth movements impact the precision of representations of phonemes in Spanish-English bilinguals?

The proposed model suggests that bilinguals have stronger encoding from visual context due to weaker phonemic encoding. This proposes that bilinguals are more likely to encode and perceive phonemes conveyed by a viseme that is presented. While monolinguals' phonemic encoding is also influenced by visual context, it is not as strong as that of bilinguals, i.e., reliance upon lip movements.

Experiment 3 addresses how selective attention, lexicosemantic encoding, and low-level sound encoding influence bilingual phonetic encoding within the model. We explore this through the phonemic restoration effect, an auditory phenomena in which listeners are able to perceptually fill-in a degraded speech signal when there is an extraneous interruption (Samuel, 1981a; Warren, 1970). The phonemic restoration effect has been found to be modulated and enhanced by both bottom-up and top-down effects (Groppe et al., 2010; Samuel, 1981a; Shahin & Miller, 2009). Top-down contextual factors such as presence of visual information, lexical identity, and sentential context have all been reported to aid in this illusory perception. Bottom-up factors, such as the type of noise/tone used and the acoustic properties of the phoneme replaced, have also been found to impact phonemic restoration.

Little research has been conducted looking at the phonemic restoration illusion in bilinguals. The goal of this experiment is to uncover the neural mechanisms underlying bilinguals' response to degraded speech and their ability, or lack of, to perceive the speech as continuous. Specifically, with the case of Spanish-English bilinguals, how does

this population respond to degraded speech in which /b/ and /v/ phonemes are removed? Furthermore, how do bilinguals perform in phonemically restoring in contexts with and without top-down input?

As seen in the proposed framework, bilinguals are more likely to give more regard to unnecessary auditory input, such as noise and the onsets of gaps within words, due to bilinguals' weaker phonemic encoding and their wider filter for low-level acoustic input (allowing stronger encoding of low-level acoustic cues). This then causes an already weak phonemic encoding to be even more deteriorated. It also suggests that due to bilinguals' weakened phonemic encoding and heavier reliance upon lexico-semantics, selective attention is necessary in filling-in missing speech.

Overall, these experiments collectively aim to understand the mechanisms of bilingual speech perception. By understanding these processes, we can better comprehend how bilinguals navigate complex listening environments and how their unique language experiences shape their phonetic perception.

Chapter 1: The McGurk illusion: A default mechanism of the auditory system

Published as Iqbal, Z.J.; Shahin, A.J.; Bortfeld, H.; Backer, K.C. The McGurk Illusion: A Default Mechanism of the Auditory System. *Brain Sci.* **2023**, *13*, 510. https://doi.org/10.3390/brainsci13030510

Abstract: Recent studies have questioned past conclusions regarding the mechanisms of the McGurk illusion, especially how McGurk susceptibility might inform our understanding of audiovisual (AV) integration. We previously proposed that the McGurk illusion is likely attributable to a default mechanism, whereby either the visual system, auditory system, or both default to specific phonemes—those implicated in the McGurk illusion. We hypothesized that the default mechanism occurs because visual stimuli with an indiscernible place of articulation (like those traditionally used in the McGurk illusion) lead to an ambiguous perceptual environment and thus a failure in AV integration. In the current study, we tested the default hypothesis as it pertains to the auditory system. Participants performed two tasks. One task was a typical McGurk illusion task, in which individuals listened to auditory-/ba/ paired with visual-/ga/ and judged what they heard. The second task was an auditory-only task, in which individuals transcribed trisyllabic words with a phoneme replaced by silence. We found that individuals' transcription of missing phonemes often defaulted to '/d/t/th/', the same phonemes often experienced during the McGurk illusion. Importantly, individuals' default rate was positively correlated with their McGurk rate. We conclude that the McGurk illusion arises when people fail to integrate visual percepts with auditory percepts, due to visual ambiguity, thus leading the auditory system to default to phonemes often implicated in the McGurk illusion.

1. Introduction

The McGurk illusion (McGurk & Macdonald, 1976) has been widely used as a model for audiovisual (AV) integration of spoken language (Beauchamp et al., 2004, 2010; Erickson et al., 2014). In the McGurk illusion, individuals exposed to audio /ba/ or /pa/ paired with a silent video of /ga/ or /ka/, respectively, often report hearing /da/ or /ta/. Based on the assumption that AV integration and the McGurk illusion rely on the same underlying neural mechanism, researchers have conducted fMRI studies using the McGurk manipulation to understand where and how AV speech integration occurs. They found heightened activity in the superior temporal sulcus/gyrus (STS/G) during McGurk perception of incongruent AV speech relative to perception of congruent speech. The STS/G was thus identified as a hub for this multisensory fusion (Beauchamp et al., 2004; Erickson et al., 2014). These findings were further corroborated via transcranial magnetic stimulation (TMS), whereupon stimulation (down regulation) of the STS with TMS (location identified via individual-specific fMRI) reduced individual susceptibility to the McGurk illusion (Beauchamp et al., 2010). In light of these and other findings, there is a widely held assumption that the McGurk illusion and AV integration are linked

mechanistically, and the McGurk illusion is used as a conduit for understanding the neurophysiology of AV development (Hirst et al., 2018; McGurk & Macdonald, 1976; Tremblay et al., 2007), AV integration as a function of aging (Sekiyama et al., 2014), and clinical deficits (e.g., schizophrenia, (Pearl et al., 2009)).

However, recent accounts from independent labs have raised doubts about past models and their conclusions regarding the mechanisms giving rise to the McGurk illusion, as well as its suitability as an index of AV integration efficacy (Getz & Toscano, 2021; Gonzales et al., 2021; Van Engen et al., 2017, 2022). Van Engen and colleagues (Van Engen et al., 2017, 2022) questioned the appropriateness of the McGurk illusion as a source of information about AV integration. Van Engen et al. (Van Engen et al., 2017), for example, showed that individuals with enhanced McGurk illusion susceptibility did not possess an added advantage in sentence recognition in noise, raising questions about the relationship between the McGurk illusion and the well-established AV integration benefit in spoken language comprehension (Grant & Seitz, 2000; Schorr et al., 2005; Sumby & Pollack, 1954). In their later study, van Engen et al. (Van Engen et al., 2022) argued that because the McGurk illusion rarely occurs naturally and is based on isolated syllables, it does not reflect real-life communication situations, thus making the McGurk illusion inappropriate to use as a procedural tool for understanding the mechanisms of AV integration in spoken language.

Moreover—and despite extensive investigation—the neural mechanisms that give rise to the McGurk illusion itself are not well understood. For example, the McGurk illusion has been distinguished from—and at times conflated with—the visual dominance illusion (Abbott & Shahin, 2018; Alsius et al., 2018; Rosenblum & Saldaña, 1992; Shahin et al., 2018), whereby individuals hear the visually conveyed phoneme when presented with an incongruent AV pair. An example of visual dominance is when exposure to audio-/ba/ paired with video-/fa/ leads to hearing /fa/, and vice versa—exposure to audio-/fa/ paired with video-/ba/ leads to hearing /ba/ (Rosenblum & Saldaña, 1992; Shahin et al., 2018). Neurophysiologically, this visual dominance effect is evident through amplitude changes in the N1-P2 auditory evoked potentials (AEPs), including the observation that the N1 AEP amplitude shifts to reflect the listener's illusory auditory perception instead of the actual auditory input (Shahin et al., 2018). It has been proposed that the McGurk illusion is a case of the visual dominance illusion (Alsius et al., 2018; Gonzales et al., 2021), partly because the McGurk illusion is reinforced when the auditory input is weakened. For example, the McGurk illusion benefits from lower sound intensity and increased noise-level (Alsius et al., 2018; G. A. Miller & Nicely, 1955). Alsius et al. (Alsius et al., 2018) argue that the McGurk illusion arises due to an array of "weak" auditory consonants, such as /b/, which can be easily confused with other stop consonants. These researchers further posit that, because place of articulation is a weak acoustic feature (Alsius et al., 2018; G. A. Miller & Nicely, 1955), the McGurk illusion is driven by vision (i.e., watching a talker's mouth movements) dominating the acoustic signal's place of articulation cues; a conclusion also hypothesized in Gonzales et al. (Gonzales et al., 2021).

Our proposal is that the McGurk illusion occurs due to ambiguity in the specific test stimuli, thus leading to a failure in AV integration, which in turn causes sensory

systems to default to specific perceptual representations. In a recent study by Gonzales et al. (Gonzales et al., 2021), we proposed that ambiguity associated with the visual stimuli due to indiscernible place of articulation (e.g., /g/, /k/, /y/), led perceivers to default to seeing '/d/t/th/' and subsequently hearing '/d/t/th/', consistent with a visual dominance account of the McGurk illusion. While the first experiment in that study supported such an account, we failed to replicate those findings in a second experiment, which featured a different talker. This lack of definitive findings motivated the current study.

Herein, we build on Gonzales et al., and further test our "auditory default" hypothesis that during the McGurk illusion, perceptual defaulting to '/d/t/th/' occurs in the auditory modality, and not in the visual modality. Participants performed two tasks. In the McGurk task, individuals saw visual-/ga/, while listening to audio-/ba/ and reported what they heard. The other task was an auditory-only task, in which individuals listened to trisyllabic words and pseudowords with one phoneme replaced by silence. We chose silence as a means to completely remove the phoneme from the acoustics and induce optimum perceptual ambiguity, a key factor in our theoretical framework. Individuals transcribed exactly what they heard as opposed to what they thought the original word/pseudoword was.

We reasoned that AV incongruency creates an ambiguous perceptual/phonetic situation, one which prevents successful AV integration and thus, allows a default auditory phonetic representation to dominate perception. Thus, we hypothesized that if the McGurk illusion is due to perceptual defaulting within the auditory modality, then (1) individuals should default to (i.e., perceptually fill-in) '/d/t/th/' (most weighted phonemes) for the silent gap in the auditory-only task; and (2) individuals with robust auditory-only '/d/t/th/' default perception should exhibit stronger McGurk susceptibility with heightened '/d/t/th/' perception. Addressing why both phenomena default to 'd/t/'th/' instead of other phonemes is beyond the scope of the present study, but we refer readers to Anderson et al., 2003 (Anderson et al., 2003) for a possible argument. Finally, we used words and pseudowords to explore how lexical knowledge influences this auditory default processes. If the '/d/t/th/' default (that is, the filling-in of missing auditory segments) is driven by prior lexical knowledge, then we should expect a higher '/d/t/th/' default in the word condition. On the other hand, we should see more '/d/t/th/' defaults in pseudowords if a '/d/t/th/' default is driven by lexical ambiguity, since pseudowords are more ambiguous than words.

2. Materials and Methods

2.1. Participants

Thirty-seven young adults participated in this study. However, three participants were excluded from data analysis due to technical issues during data collection or not being a native/fluent English speaker, resulting in usable data from thirty-four participants (>18 years of age, M = 20.73 years, SD = 2.13 years, 2 participants did not provide their specific age; 27 females, 4 males, 3 did not respond; native or fluent English speakers). Of these thirty-four individuals, 29 reported that they are right-handed, 2 left-handed, and 3 ambidextrous. All participants self-reported normal hearing, normal

or corrected vision, and no language deficits. Participants were recruited via an internal recruiting system of the University of California, Merced and provided written consent prior to participation. All experimental protocols were approved by the Institutional Review Board (IRB) of the University of California, Merced, and all methods were carried out in accordance with the guidelines and regulations of the IRB of the University of California, Merced and in accordance with the Declaration of Helsinki. Participants were monetarily compensated for their participation.

2.2. Stimuli

The study consisted of two tasks. The stimuli in one task consisted of English words and pseudowords spoken by a female talker (mean f0 = 203 Hz; see Shahin et al., 2009 & Shahin & Miller, 2009 for more details) with one phoneme replaced by silence. There was a total of 39 words and 36 pseudowords. Consonants were manually removed in Adobe Audition (Adobe Systems Inc., San Jose, CA, USA). For each word/pseudoword, one of the following consonants was removed from either the second or third syllable and replaced with silence: /k, t, d, g, b, ʃ, s, ʒ, z, ʧ, ʤ, l, r/. These phonemes were selected to ensure that there was a distribution of 2–4 consonants for each category of manner or place of articulation (i.e., stops, fricatives, bilabial, alveolar, etc.). Furthermore, the number of consonants removed from the second or third syllable was balanced across words and pseudowords; 15 consonants and 14 consonants were removed from the second syllable of words and pseudowords, respectively, and 24 and 22 consonants were removed from the third syllable of words and pseudowords, respectively.

The other task involved a classic McGurk design whereby individuals listened to and watched a talker (two female talkers, mean f0 = 199 Hz, 184 Hz) uttering congruent and incongruent consonant vowels (CVs). The purpose of this task was to test the subjects for McGurk susceptibility. To create the stimuli, audio recordings of /ba/, /da/, and /ga/ were used along with video recordings of the talkers producing the same CVs. The videos were cropped, ensuring that participants could only view the space between the bridge of the talker's nose and the bottom of the neck. This was to encourage participants to focus on the mouth and not be distracted by other features, such as the talker's eyes. To create the AV pairings of congruent /ba/, congruent /da/, congruent /ga/, and incongruent /ba-ga/ (auditory /ba/ paired with visual /ga/ or viseme /ga/), the auditory stimuli of each talker were temporally aligned to the acoustic onset of the video, respectively. The temporal alignment included a natural auditory delay as is typical of natural utterances. This resulted in 32 stimuli (2 talkers × 4 AV pairings × 4 exemplars). The first half of participants were presented with stimuli of the first talker and the remaining participants were presented with the second talker.

2.3. Procedure

Participants sat in an enclosed room about 90 cm from a 27-inch computer monitor with two external speakers on either side, located at a 45-degree angle relative to the listener. Participants were given two tasks: the word/pseudoword task, which was split into 2 blocks for each set of stimuli, and the AV McGurk task (1 block). This totaled

3 blocks across the 2 tasks; the order of the tasks was counterbalanced across participants.

Prior to the start of the Word and Pseudoword blocks, participants were told they would be presented with words, which might sound like real English words or might seem similar to English words. Their task was to listen carefully and type out to the best of their ability exactly what they heard, and not what they lexically thought they heard. These instructions were repeated once again on the monitor prior to the start of the experiment. Participants typed their responses using a keyboard. Stimuli were presented using Presentation v.20.3 (Neurobehavioral Systems, Inc., Berkeley, CA, USA). The Word and Pseudoword blocks consisted of either 39 or 36 trials, respectively, with each stimulus played only once.

For the AV block, prior to starting, participants were informed that they would be presented with videos of an individual producing speech sounds. It was emphasized that participants should always be paying attention to the screen to ensure that they were focused on the talker's mouth movements. The participants' task was to type out what they heard. If they heard an ambiguous percept, they were told to transcribe the most dominant percept. These instructions were also presented on the monitor prior to the start of the block. There was a total of 32 trials, with each stimulus repeated twice. An optional two-minute break was offered to participants between each block to mitigate boredom and fatigue.

2.4. Data Analysis

Logfiles of participants' responses were transferred to Excel spreadsheets, which were then parsed using in-house custom MATLAB code (MathWorks, Natick, MA, USA). For the Word and Pseudoword blocks, responses were extracted for each word or pseudoword for each participant. The output of this parsing code was a table containing information about the stimulus, which phoneme was removed in the stimulus, and the response for each trial. An additional column was manually completed, in which we recorded how a participant perceived the removed phoneme on that trial. For example, if the word presented was "addition" with the /ʃ/ ('sh' sound) removed and a participant reported perceiving "addithen", this was coded as 'th' filling-in. There were instances where participants perceived no change at all from the original word before a phoneme was removed (e.g., perceiving "addition") and cases where they reported perceiving the gap itself (e.g., perceiving "addi _on").

For the AV block, responses were categorized according to the first letter transcribed by the participant (i.e., responses "ba", "bah", and "bo" were all included in the response category /b/). The output was a table containing information about the auditory token, visual token, and the first-letter response for each trial.

2.5. Statistical Analysis

The data were statistically analyzed in R (R Core Team, 2019) and MATLAB. Two types of analyses were performed on the data: (1) Mixed effect multinomial logistic

regression analyses performed on the auditory-only task data (conducted in R (R Core Team, 2019), and (2) a correlation analysis to examine the relationship between '/d/t/th/' perception on the auditory-only task and on the McGurk trials (conducted in MATLAB).

First, we conducted a mixed effects multinomial logistic regression using the mclogit package (Elff, 2022), to examine whether the Block Type (Words vs. Pseudowords) may predict subjects' auditory perception of the missing phoneme. The outcome measure was the perception of the silent gap (i.e., Response), which comprised four categories: '/d/t/th/' (i.e., filled in the silent gap incorrectly with /d/,/t/, or /th/), Gap (i.e., perceived the silence as a gap without any phonetic filling-in), No Change (i.e., perceived the word or pseudoword by filling in the silent gap with the correct phoneme), and Other Phoneme (i.e., filled in the silent gap with any phoneme except for/d/,/t/,/th/, or the correct phoneme). The /d/t/th/ Response category was set as the referent level, since it was the variable of interest that we wanted to contrast with the other three categories. Importantly, the contrast between /d/t/th/ and Other Phoneme was done to test the hypothesis that when individuals fill-in the silent gap with an incorrect phoneme, they should perceive '/d/t/th/' more often than the other phonemes—especially for the Pseudoword stimuli.

This initial model included only the fixed effect of Block Type (reference level: Pseudowords), as well as the intercept corresponding to each subject as a random effect. The formula was Response ~ Block Type + 1|SubjectID. An effect of Block Type would reveal that lexical context drives perception of the missing phoneme. This mixed effects multinomial logistic regression model was run using the mblogit function, with the method for modeling the random effects set to the Penalized Quasi-Likelihood (PQL) method. Single-trial data, totaling 2537 trials across all 34 subjects, were inputted into the model with 13 trials (i.e., 0.5% of all 2550 trials) excluded due to missing responses (i.e., the subject pressed "Enter" without typing anything). Relative Risk Ratios (RRR) were computed by exponentiating the coefficients for the fixed effects. In the context of the current analysis, an RRR greater than 1 indicates that missing phonemes in Words are more likely than missing phonemes in Pseudowords to be perceived as Other Phoneme (or Gap or No Change) over '/d/t/th/'. An RRR less than 1 indicates the opposite pattern, for example, relative to missing phonemes within Pseudowords, participants were more likely to perceive missing phonemes within Words as '/d/t/th/' than the contrasted perceptual outcome (i.e., Other Phoneme, No Change, or Gap).

We also conducted a secondary mixed-effects multinomial logistic regression analysis, which was an exploratory analysis to assess whether the Manner of Articulation (MoA) of the missing phoneme, as well as its interaction with Block Type, predict auditory perception of the missing phoneme. In this second model, the fixed effects included the Block Type (2 levels: Word or Pseudoword), Manner of Articulation (MoA) of the missing phoneme (3 levels: Fricative/Affricate [ʃ, s, ʒ, z, ʧ, ʤ], Liquid [l, r], and Stop [k, t, d, g, b]), and their interaction, as well as the Syllable from which the phoneme was deleted (2 levels: 2nd or 3rd Syllable). For the fixed effects, the reference levels for the three predictors were the Pseudoword Block Type, the Fricative MoA, and the 2nd Syllable, respectively. Syllable was inputted as a fixed effect to control for any possible syllable effects on perception, since the syllable from which the missing phoneme was

removed was not balanced across the MoA categories, as follows: A fricative/affricate was removed from the 2nd syllable of 3 words and 4 pseudowords, and from the 3rd syllable of 20 words and 15 pseudowords. A liquid was removed from the 2nd syllable of 4 words and 4 pseudowords, and from the 3rd syllable of 2 words and 2 pseudowords. A stop consonant was removed from the 2nd syllable of 8 words and 6 pseudowords, and from the 3rd syllable of 2 words and 5 pseudowords. However, the effects of Block Type and MoA, as well as their interaction, were the key effects of interest. The intercept corresponding to each subject was inputted into the model as a random effect. The formula was Response ~ Block Type * MoA + Syllable + 1|SubjectID. To foreshadow the results, follow-up multinomial logistic regression analyses were conducted to interpret significant interaction effects.

Finally, the correlation analysis was done to test the main hypothesis that individuals with stronger auditory-only '/d/t/th/' perception of missing phonemes should exhibit increased '/d/t/th/' perception of the McGurk stimuli. To do this, the total percentage of '/d/t/th/' responses for the auditory-only filling-in trials (collapsed across the type of missing phoneme and word/pseudoword condition) was computed for each subject. Similarly, the total percentage of '/d/t/th/' responses on the McGurk trials was computed for each subject. Subsequently, we conducted Pearson correlations on these two sets of '/d/t/th/' percentages.

3. Results

3.1. Mixed Effects Multinomial Logistic Regression

The primary multinomial logistic regression analysis was performed to examine how the presence (or absence) of lexico-semantic context affects perception of a missing phoneme in auditory-only stimuli. This was done by inputting Block Type (Words vs. Pseudowords) as a fixed effect into the model. The results are depicted in Table 1 and Figure 1. Recall that '/d/t/th/' perception was set as the referent level for the outcome measure, so that it could be compared with the other three percept categories (Gap, No Change, and Other Phoneme). The relative risk of perceiving the missing phoneme as a Gap vs. '/d/t/th/', No Change vs. '/d/t/th/', and Other Phoneme vs. '/d/t/th/' for Words was significantly higher (i.e., 1.53, 3.86, and 1.41 times higher, respectively), than the same relative risks for Pseudowords. Thus, the presence of lexico-semantic context significantly affected perception, such that participants were most likely to correctly fill-in the missing phoneme for the Word stimuli. On the flip side, missing phonemes within Pseudowords were significantly more likely to be perceived as '/d/t/th/' than both the Gap and Other Phoneme percepts.

*Table 1: Results of the Multinomial Logistic Regression to examine the effect of Block Type (Words vs. Pseudowords) on perception. Significant fixed effects are depicted in bold font and indicated with asterisks as follows: ** $p < 0.01$, *** $p < 0.001$.*

Fixed Effects						
Contrast	**Effect**	**RRR**	**95% CI (LL)**	**95% CI (UL)**	**z**	***p***

Gap vs. /d/t/th/	**Intercept**	**0.30**	**0.19**	**0.47**	**−5.21**	**<0.001 *** **
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**1.53**	**1.14**	**2.04**	**2.86**	**0.004 ** **
No Change vs. /d/t/th/	Intercept	0.88	0.69	1.11	−1.09	0.276
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**3.86**	**3.14**	**4.73**	**12.89**	**<0.001 *** **
Other Phoneme vs. /d/t/th/	**Intercept**	**0.49**	**0.39**	**0.63**	**−5.56**	**<0.001 *** **
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**1.41**	**1.09**	**1.83**	**2.62**	**0.009 ** **

Random Effects

Intercept (Subject ID) Co-variance Parameters	Gap~1		NoChange~1		Other~1	
	Estimate	SE	Estimate	SE	Estimate	SE
Gap~1	1.47	1.21				
No Change~1	0.01	0.08	0.31	0.01		
Other~1	0.26	0.26	0.07	0.02	0.30	0.06

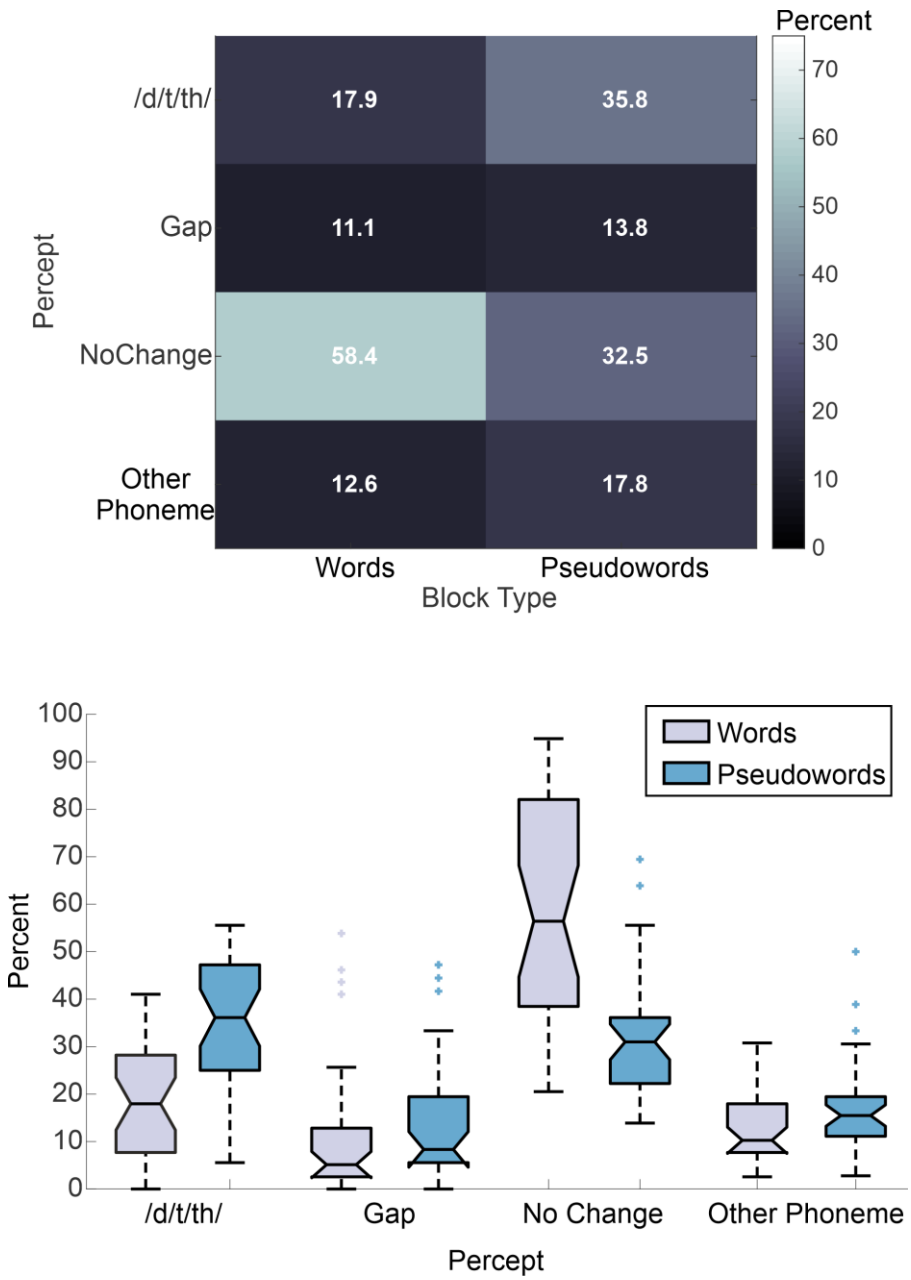

Figure 1: Results of the Auditory-Only Filling-in Task. Top portion shows the group average percentages of each percept, separately for the Word and Pseudoword stimuli. Bottom portion shows box plots of the same data.

Upon close examination of the data, it appeared that the missing phoneme's Manner of Articulation (MoA) may modulate auditory perception. Thus, we also ran a more complex, follow-up multinomial logistic regression as an exploratory analysis to examine whether the Manner of Articulation (MoA) of the missing phoneme and its interaction with Block Type, predict auditory perception of the missing phoneme, while controlling for the Syllable from which the missing phoneme was removed. The results of this mixed effects multinomial regression are presented in Table 2. Additionally, Figure 2

illustrates the percentages that each percept experienced across subjects, within each Block Type (Words, Pseudowords) and MoA.

*Table 2. Results of the Multinomial Logistic Regression to examine the interaction between Block Type and Manner of Articulation on auditory perception (while controlling for the effect of Syllable). Significant fixed effects/interactions are depicted in bold font and indicated with asterisks as follows: ** p < 0.01, *** p < 0.001.*

Fixed Effects						
Contrast	**Effect**	**RRR**	**95% CI (LL)**	**95% CI (UL)**	**z**	***p***
Gap vs. /d/t/th/	**Intercept**	**0.19**	**0.11**	**0.35**	**−5.54**	**<0.001 *** **
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**1.81**	**1.26**	**2.59**	**3.22**	**0.001 ** **
	MoA-Fricative (ref.)					
	MoA-Liquid	**11.12**	**6.02**	**20.54**	**7.69**	**<0.001 *** **
	MoA-Stop	**2.12**	**1.24**	**3.65**	**2.73**	**0.006 ** **
	Syllable-2 (ref.)					
	Syllable-3	0.92	0.64	1.32	−0.45	0.654
	Block Type-Word * MoA-Liquid	1.53	0.30	7.71	0.52	0.606
	Block Type-Word * MoA-Stop	-	-	-	-	-
No Change vs. /d/t/th/	**Intercept**	**0.14**	**0.09**	**0.21**	**−9.00**	**<0.001 *** **
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**4.43**	**3.35**	**5.87**	**10.42**	**<0.001 *** **
	MoA-Fricative(ref.)					
	MoA-Liquid	**15.66**	**8.89**	**27.60**	**9.52**	**<0.001 *** **
	MoA-Stop	**16.83**	**11.28**	**25.10**	**13.83**	**<0.001 *** **
	Syllable-2 (ref.)					
	Syllable-3	**2.69**	**1.99**	**3.64**	**6.48**	**<0.001 *** **
	Block Type-Word * MoA-Liquid	**7.39**	**1.63**	**33.51**	**2.60**	**0.009 ** **
	Block Type-Word * MoA-Stop	-	-	-	-	-
Other Phoneme vs. /d/t/th/	**Intercept**	**0.20**	**0.13**	**0.30**	**−7.73**	**<0.001 *** **
	Block Type-Pseudoword (ref.)					
	Block Type-Word	**2.06**	**1.46**	**2.90**	**4.15**	**<0.001 *** **
	MoA-Fricative(ref.)					

	MoA-Liquid	**15.00**	**8.42**	**26.71**	**9.20**	**<0.001 *****
	MoA-Stop	**6.35**	**4.05**	**9.95**	**8.06**	**<0.001 *****
	Syllable-2 (ref.)					
	Syllable-3	1.16	0.84	1.62	0.91	0.364
	Block Type-Word * MoA-Liquid	2.50	0.53	11.83	1.15	0.249
	Block Type-Word * MoA-Stop	-	-	-	-	-
Random Effects						
Intercept (Subject ID) Co-variance Parameters	Gap ~1		No Change ~1		Other ~1	
	Estimate	SE	Estimate	SE	Estimate	SE
Gap ~1	1.72	2.63				
No Change ~1	0.20	0.65	0.63	0.19		
Other ~1	0.37	0.76	0.22	0.20	0.38	0.23

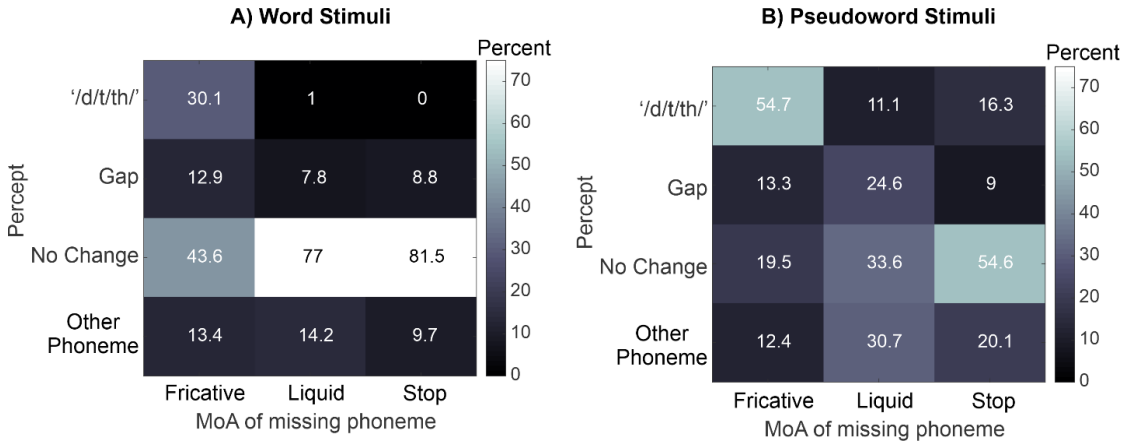

Figure 2. Group-average percentages of each percept reported for the (A) Word and (B) Pseudoword stimuli, depending on the MoA of the missing phoneme. The percentages of each percept were calculated separately within each MoA category (i.e., each column adds up to 100%).

As shown in Table 2, there was a significant interaction between Block Type and MoA for the No Change vs. '/d/t/th/' contrast. Thus, a follow-up mixed effects multinomial logistic regression was performed to facilitate interpretation of the results. The follow-up analysis was done by re-coding the two predictors, Block Type and MoA, into a single predictor variable, called "BTMoA" which had six levels encoding both the Block Type (Word, Pseudoword) and the Manner of Articulation of the missing phoneme (i.e., Word-Fricative, Word-Stop, Word-Liquid, Pseudoword-Fricative, Pseudoword-Stop, and Pseudoword-Liquid). Like the initial model, the '/d/t/th/' response was set as the referent level for the outcome measure. For the fixed effects, the reference levels included the Pseudoword-Fricative condition and the 2nd Syllable. The intercept corresponding to each subject was inputted into the model as a random effect. The formula for the follow-up model was Response ~ BTMoA + Syllable + 1|SubjectID. An identical follow-up analysis was conducted, but with Word-Fricative as the reference level to directly contrast the effect of MoA within the Word block type.

The results of these follow-up analyses are depicted in Tables 3 and 4. Please note that a valid estimate could not be generated for the Word*MoA-Stop interaction in the initial multinomial regression and for the Word-Stop condition in these follow-up analyses. Close examination of the data revealed that across the 34 participants, there was never a Word-Stop trial in which the missing phoneme was incorrectly perceived as /d/t/th/; thus, a valid estimate could not be generated. (Word-Stop and Pseudoword-Stop stimuli with a missing /d/ or /t/ that was subsequently perceived as /d/ or /t/, respectively, were categorized as No Change.).

*Table 3. Results of the follow-up mixed effects multinomial logistic regression, with Pseudoword-Fricative as the reference category for the Block Type-Manner of Articulation condition. Significant fixed effects/interactions are depicted in bold font and indicated with asterisks as follows: ** $p < 0.01$, *** $p < 0.001$.*

Fixed Effects						
Contrast	**Effect**	**RRR**	**95% CI (LL)**	**95% CI (UL)**	**z**	***p***
Gap vs. /d/t/th/	**Intercept**	**0.19**	**0.11**	**0.35**	**−5.56**	**<0.001 *****
	Pseudoword-Fricative (ref.)					
	Pseudoword-Liquid	**11.10**	**6.01**	**20.51**	**7.69**	**<0.001 *****
	Pseudoword-Stop	**2.12**	**1.23**	**3.64**	**2.72**	**0.006 ****
	Word-Fricative	**1.81**	**1.26**	**2.59**	**3.22**	**0.001 ****
	Word-Liquid	**30.69**	**6.69**	**140.67**	**4.41**	**<0.001 *****
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					
	Syllable-3	0.92	0.64	1.32	−0.45	0.653
No Change vs. /d/t/th/	**Intercept**	**0.14**	**0.09**	**0.21**	**−9.01**	**<0.001 *****
	Pseudoword-Fricative (ref.)					
	Pseudoword-Liquid	**15.65**	**8.88**	**27.58**	**9.51**	**<0.001 *****
	Pseudoword-Stop	**16.82**	**11.27**	**25.09**	**13.83**	**<0.001 *****
	Word-Fricative	**4.43**	**3.35**	**5.87**	**10.42**	**<0.001 *****
	Word-Liquid	**513.19**	**122.58**	**2148.45**	**8.54**	**<0.001 *****
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					
	Syllable-3	**2.69**	**2.00**	**3.63**	**6.45**	**<0.001 *****
Other Phoneme vs. /d/t/th/	**Intercept**	**0.20**	**0.13**	**0.30**	**−7.74**	**<0.001 *****
	Pseudoword-Fricative(ref.)					
	Pseudoword-Liquid	**14.99**	**8.42**	**26.69**	**9.20**	**<0.001 *****
	Pseudoword-Stop	**6.35**	**4.05**	**9.95**	**8.06**	**<0.001 *****
	Word-Fricative	**2.06**	**1.46**	**2.90**	**4.15**	**<0.001 *****
	Word-Liquid	**77.24**	**17.76**	**335.97**	**5.80**	**<0.001 *****
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					

	Syllable-3	1.16	0.84	1.62	0.91	0.364
Random Effects						
Intercept (Subject ID) Co-variance Parameters	Gap ~1		No Change ~1		Other ~1	
	Estimate	SE	Estimate	SE	Estimate	SE
Gap ~1	1.69	2.44				
No Change ~1	0.19	0.59	0.63	0.18		
Other ~1	0.36	0.69	0.22	0.18	0.38	0.21

*Table 4. Results of the follow-up mixed effects multinomial logistic regression, with Word-Fricative as the reference category for the Block Type-Manner of Articulation condition. Please note that the random effects covariance parameters are not displayed, since they are identical to Table 3. Significant fixed effects/interactions are depicted in bold font and indicated with asterisks as follows: * p < 0.05, ** p < 0.01, *** p < 0.001.*

Fixed Effects						
Contrast	**Effect**	**RRR**	**95% CI (LL)**	**95% CI (UL)**	**z**	***p***
Gap vs. /d/t/th/	**Intercept**	**0.35**	**0.19**	**0.63**	**−3.50**	**<0.001 *** **
	Word-Fricative(ref.)					
	Pseudoword-Fricative	**0.55**	**0.39**	**0.79**	**−3.21**	**0.001 ** **
	Pseudoword-Liquid	**6.15**	**3.32**	**11.40**	**5.76**	**<0.001 *** **
	Pseudoword-Stop	1.17	0.68	2.02	0.58	0.563
	Word-Liquid	**17.00**	**3.70**	**78.01**	**3.64**	**<0.001 *** **
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					
	Syllable-3	0.92	0.64	1.32	−0.45	0.653
No Change vs. /d/t/th/	**Intercept**	**0.61**	**0.40**	**0.93**	**−2.32**	**0.020 * **
	Word-Fricative(ref.)					
	Pseudoword-Fricative	**0.23**	**0.17**	**0.30**	**−10.42**	**<0.001 *** **
	Pseudoword-Liquid	**3.53**	**2.03**	**6.14**	**4.47**	**<0.001 *** **
	Pseudoword-Stop	**3.79**	**2.60**	**5.54**	**6.90**	**<0.001 *** **
	Word-Liquid	**115.76**	**27.83**	**481.56**	**6.53**	**<0.001 *** **
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					
	Syllable-3	**2.69**	**1.99**	**3.64**	**6.45**	**<0.001 *** **
Other Phoneme vs. /d/t/th/	**Intercept**	**0.40**	**0.26**	**0.61**	**−4.27**	**<0.001 *** **
	Word- Fricative(ref.)					
	Pseudoword-Fricative	**0.23**	**0.17**	**0.30**	**−4.14**	**<0.001 *** **
	Pseudoword-Liquid	**7.27**	**4.09**	**12.93**	**6.76**	**<0.001 *** **
	Pseudoword-Stop	**3.08**	**1.97**	**4.81**	**4.94**	**<0.001 *** **
	Word-Liquid	**37.47**	**8.62**	**162.87**	**4.83**	**<0.001 *** **
	Word-Stop	-	-	-	-	-
	Syllable-2 (ref.)					

	Syllable-3	1.16	0.84	1.62	0.91	0.364

As shown in Table 3, while controlling for the syllable containing the missing phoneme, the relative risks of perceiving Gap vs. /d/t/th/ for the Pseudoword-Liquid, Pseudoword-Stop, Word-Fricative, and Word-Liquid conditions were significantly higher than the same relative risk for the Pseudoword-Fricative condition. The same pattern of results was observed for the No Change vs. /d/t/th/ and Other Phoneme vs. /d/t/th/ contrasts. Specifically, when the Pseudoword-Fricative condition was set as the reference level, the relative risk of perceiving a Gap, No Change, or Other Phoneme vs. /d/t/th/ was 1.81, 4.43, and 2.06 times higher, respectively, for the Word-Fricative condition. Overall, the Pseudoword-Fricative condition was most likely to lead to /d/t/th/ perception compared to the other conditions. As demonstrated in Table 4, while controlling for the syllable with the missing phoneme, the relative risks of perceiving Gap vs. /d/t/th/ for the Pseudoword-Liquid and Word-Liquid conditions were significantly higher than that for the Word-Fricative condition. Moreover, the relative risks of perceiving No Change or Other Phoneme vs. /d/t/th/ for the Pseudoword-Liquid, Pseudoword-Stop, and Word-Liquid conditions were also significantly greater than that for the Word-Fricative condition.

Taken together, these results suggest that participants were most likely to perceive /d/t/th/ in the place of a missing fricative, and even to a greater extent when the fricative was deleted from a Pseudoword compared to a Word. When the missing phoneme was a liquid, participants were most likely to perceive No Change if the stimulus was a Word, or most likely to perceive a Gap or Other Phoneme if the stimulus was a Pseudoword. Therefore, both the lexical and articulatory context (and their interaction) seem to play a role in shaping listeners' perception of missing phonemes.

Finally, another incidental finding of this analysis was an effect of Syllable. Specifically, when the missing phoneme was removed from the third syllable, the relative risk of perceiving No Change vs. /d/t/th/ was significantly higher (RRR = 2.69, $p < 0.001$) than when the missing phoneme was removed from the second syllable. There was no effect of Syllable for the Gap vs. /d/t/th/or Other Phoneme vs. /d/t/th/ contrasts. This again suggests an effect of context, such that increasing the amount of preceding context within these trisyllabic words and pseudowords facilitated accurate filling-in of the missing phoneme.

3.2. Correlation between Auditory-Only Filling-In and McGurk Illusion

On the audiovisual task, participants performed well on the congruent audiovisual trials, which included congruent /ba/ (96.3% ± 2.1% [mean ± se]), congruent /ga/ (100% ± 0%), and congruent /da/ (89.0% ± 3.5%) stimuli. On the incongruent audiovisual (McGurk) trials comprising /ba/-/ga/ stimuli, there was wide variability in subjects' susceptibility to the McGurk illusion; 16 of the 34 subjects never experienced the McGurk illusion (i.e., never perceived /d/,/t/, or /th/ on any of the incongruent trials), and 3 of the 34 subjects experienced the illusion on 100% of the incongruent trials.

Crucially, as shown in Figure 3, there was a significant across-subjects correlation between the percentage of '/d/t/th/' perception on the auditory-only task (collapsed across word and pseudoword trials) and on the McGurk trials (r = 0.397, p = 0.020). Subjects who perceived the McGurk illusion more often also tended to perceive '/d/t/th/' more often on the auditory-only trials. We also conducted follow-up correlations for the word and pseudoword trials separately. The correlation for the pseudoword trials was significant (r = 0.386, p = 0.024), and the correlation for the word trials was marginally significant (r = 0.306, p = 0.078).

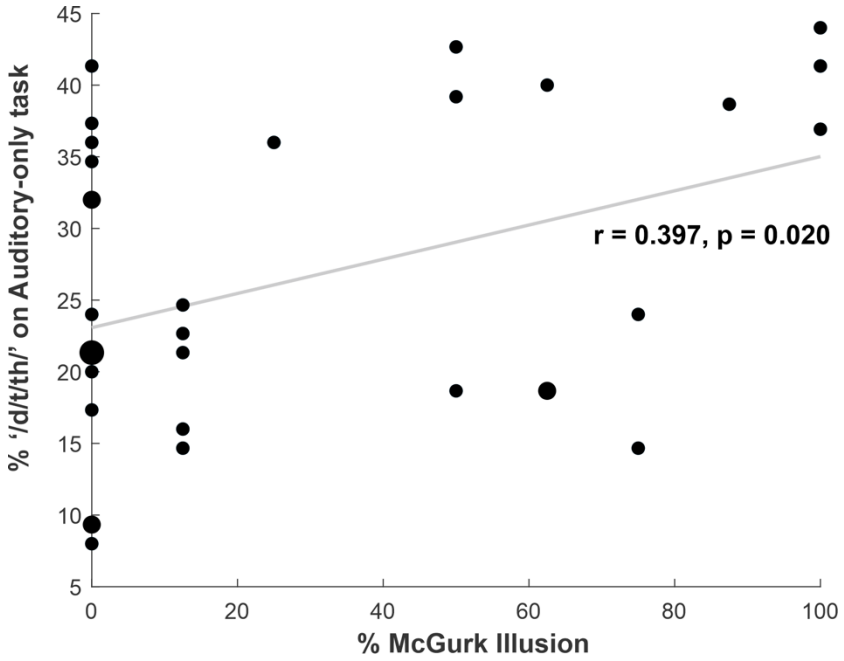

Figure 3. Results of the correlation between /d/t/th/ perception on McGurk trials and on the Auditory-only task. The dots represent individual participants; enlarged dots depict data points shared by two or more participants, with dot size proportional to the number of overlapping participants.

4. Discussion

Our results point to an auditory default mechanism whereby AV integration fails due to ambiguity in the visual stimuli, forcing the auditory modality to drive the McGurk illusion on its own (i.e., auditory default). Notably, the results showed that listeners often perceive a gap in auditory-only stimuli as '/d/t/th/'—the same percept often perceived during the McGurk illusion. Furthermore, participants who were more likely to perceive '/d/t/th/' on the auditory-only task were also more likely to perceive '/d/t/th/' illusory perception on the McGurk task.

A remaining challenge is to identify the neuronal basis for this proposed auditory default mechanism. We begin by arguing that current evidence calls into question links between AV illusions and AV integration. We conclude by proposing that the current

results support the existence of a default mechanism that favors the auditory modality and thus gives rise to the classic McGurk illusion.

First, assuming parsimony, the same general AV mechanism should underlie the classic McGurk and visual dominance illusions alike, even though they are induced by different pairings of visual and auditory syllables. Just because the stimuli are different, it does not necessarily follow that the AV processing mechanism is distinct. Otherwise, AV processing would be highly inefficient. Second, we know that the general influence of visual modality on auditory modality is suppressive (Besle et al., 2004; Pilling, 2009; Shatzer et al., 2018; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005), a finding further confirmed in our own lab (Shahin et al., 2018). Third, there is also evidence for a secondary influence of the visual modality on the auditory modality: an encoding phase, in which the visual modality encodes its phonetic representation (viseme) within the auditory modality (Abbott & Shahin, 2018; Shahin et al., 2018; Smith et al., 2013). Indeed, in Shahin et al. (Shahin et al., 2018), we demonstrated that visual suppression of the auditory cortex mentioned in the second premise above is deliberate. This cross-modal suppression occurs so that existing auditory representations conveyed by the ear are inhibited to render the auditory modality more prone to alteration by the visual modality (i.e., the cross-modal secondary encoding phase; third premise above). This is necessary because if auditory representations are too robust, it would be difficult for the visual modality to overwrite them. Shahin et al. (Shahin et al., 2018) used the visual dominance illusion to demonstrate this effect: while the N1-P2 auditory evoked potentials were suppressed for AV versus auditory-only conditions; there was a specific encoding effect as well. When individuals heard ‘ba’ when presented with visual-/ba/ and auditory-/fa/, the auditory N1 increased in amplitude (i.e., became more negative). When individuals heard ‘fa’ when presented with visual-/fa/ and auditory-/ba/, the auditory N1 decreased in amplitude. This shift mirrored the relative amplitude difference for /ba/ and /fa/ in the auditory-only condition, with the N1 for /ba/ being larger (more negative) than the N1 for /fa/.

Based on the above, we propose a tentative model of the McGurk illusion mechanism. Our theoretical framework illustrated in Figure 4, posits that the classic McGurk illusion follows the same process as the visual dominance illusion, except for one step. Following inhibition of phonetic representations within the auditory modality, the encoding step fails to materialize because the visual utterance of /ga/ or /ka/ is indiscernible—it is confused with /sa/, /ya/, /ha/, /ja/ (Gonzales et al., 2021). Consequently, the auditory system is faced with an ambiguous situation: the auditory input has been inhibited while there is no discernable visual input. As a result, the auditory modality is forced to default to phonetic representations that are naturally dominant (highly weighted) in discourse (i.e., ‘/d/t/th/’). Our framework is consistent with animal work (Ghazanfar, 2005; Kayser et al., 2008, 2010), but deviates from other models, which assert that decisions about multisensory integration occur in higher-level brain regions, such as superior temporal sulcus/gyrus and/or prefrontal cortex (Beauchamp et al., 2004; Calvert et al., 2000; Hwang & Romanski, 2015; Noppeney et al., 2010; Romanski, 2012). These high-level networks evaluate the sum of visual and auditory input and conclude upon a percept (Erickson et al., 2014; L. M. Miller & D’Esposito, 2005; Morís Fernández et al., 2017).

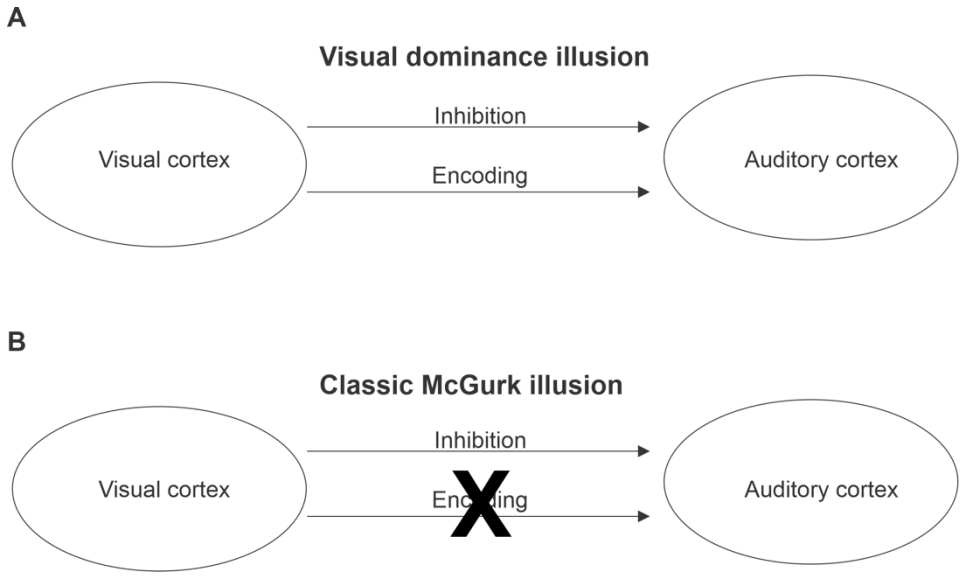

*Figure 4. Theoretical Framework illustrating underlying mechanism of the (**A**) Visual dominance illusion and (**B**) Classic McGurk illusion. The Visual Dominance illusion involves visually-mediated inhibition of the auditory cortex, followed by encoding of the visually-conveyed information at the auditory cortex, leading to auditory perception of the visually-conveyed phoneme. According to the proposed "Auditory Default" mechanism underlying the Classic McGurk Illusion, visually-mediated inhibition of the auditory cortex occurs, but because the visual phonetic representation is ambiguous, the visually-mediated encoding step fails, and thus auditory perception is dominated by the phonetic representations with the intrinsically strongest weights within the auditory cortex ('/d/t/th/').*

As for lexical influence, it is interesting that in the auditory-only task, individuals defaulted to 'd/t/th/' more often when hearing pseudowords than words. This lexical effect is not surprising given what we know about the phonemic restoration (PR) phenomenon (also known as illusory filling-in). In PR, words with noise-replaced segments can be heard as continuing through the noise (i.e., the speech is perceived as intact) (Samuel, 1981a, 1981b; Shahin et al., 2009; Shahin & Miller, 2009; Warren, 1970; Warren et al., 1997). The primary difference between the auditory-only task and PR is the difference in replaced segments—silence versus noise. Words exhibit stronger PR than pseudowords, and the more syllables that are within the word, the more robust the PR illusion (Samuel, 1981a); both of these effects are consistent with the results from the present auditory-only filling-in task. Moreover, noise is stronger at eliciting PR than silence (Warren & Obusek, 1971). However, an interesting result is that insertion of a small silent gap coupled with the noise enhances restoration of stop consonants (Samuel, 1981b; Sherman, 1971; Warren & Obusek, 1971). Indeed, even in the present study, relative to pseudowords or words with a missing fricative/affricate, pseudowords with a stop consonant replaced completely by silence were more likely to be accurately filled-in than perceived as '/d/t/th/', and words with a missing stop consonant were never incorrectly perceived as '/d/t/th/'. Interestingly, words and pseudowords with missing

fricatives/affricates were most often filled-in incorrectly with '/d/t/th/'. Together, these results suggest that the lexical and articulatory context modulate the auditory-only filling-in process, but further research is needed to fully understand these incidental findings as they are beyond the scope of the present manuscript.

There are a few additional issues that warrant our attention. First, recent reports have raised doubts about the suitability of the McGurk illusion as a tool for understanding AV integration of spoken language (Van Engen et al., 2017, 2022). These researchers assert that the McGurk illusion is based on stimulus manipulations that are rare in real life situations (in particular, see (Van Engen et al., 2022)). Such a view is consistent with the reasoning laid out in a recent review of visual-only illusions (Rogers, 2022). However, we do not subscribe to this reasoning. Our view is that manipulating stimuli in ways that rarely—or never—occur in real-life is a powerful way to understand the mechanisms underlying real-life processing. After all, it is our rich history of experiencing situations in real-life that likely gives rise to the experience of illusions. Thus, reverse investigation ("reverse engineering") is key to understanding the mechanisms at play in ecologically valid situations. Second, because our study suggests that the McGurk illusion is a consequence of failure to integrate AV percepts, it is not surprising that individuals who experience this illusion do not perform better on ecologically valid AV speech comprehension tasks, in line with the conclusions of van Engen et al. (Van Engen et al., 2017). Third, while incongruent AV stimuli in spoken language rarely existed decades ago, we now encounter them often in communication due to video conferencing. In a way, the discovery of the McGurk illusion in 1976, has significantly impacted our understanding of an evolving perceptual phenomenon that is currently often encountered in real life situations—audiovisual incongruency. Thus, the McGurk illusion was well ahead of its time.

5. Conclusions

In the current study, we argue that the well-known McGurk illusion may arise due to a failure of audiovisual integration. Consequently, perception is exclusively determined within the auditory modality, such that perception favors (i.e., defaults to) the phonemes often implicated in the McGurk illusion, /d/t/th/. For these reasons, the McGurk illusion is not well suited as a tool to study AV mechanisms in spoken language. However, the McGurk illusion remains an outstanding discovery in language perception research, one that has significantly advanced knowledge in the field.

Author Contributions: Conceptualization, Z.J.I., A.J.S., H.B., and K.C.B.; methodology, Z.J.I., A.J.S., H.B., and K.C.B.; software, Z.J.I. and K.C.B.; formal analysis, Z.J.I., A.J.S., H.B., and K.C.B.; investigation, Z.J.I.; writing—original draft preparation, Z.J.I., A.J.S., and K.C.B.; writing—review and editing, Z.J.I., A.J.S., H.B., and K.C.B.; visualization, Z.J.I., A.J.S., H.B., and K.C.B.; supervision, A.J.S. and K.C.B.; funding acquisition, A.J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Institute on Deafness and Other Communications Disorders Grant R01 DC-013543 (A.J.S.).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of California, Merced (protocol code UCM2019-4, approved 10 July 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: To access the data collected from this experiment, please click on the link to the Figshare site: https://figshare.com/projects/The_McGurk_illusion_A_default_mechanism_of_the_auditory_system/156281

Acknowledgments: The authors thank Madison Lacanlale for assistance with data collection, and Yueqi Yan for guidance on the statistical analyses.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Chapter 2: The effects of visual context on speech perception across language dominance

1. Introduction

In our everyday lives, we experience a constant stream of incoming information arriving from our multiple senses. In cases where these multisensory inputs likely originate from the same source, our brain does its best to integrate them into one percept. One of the most profound multisensory processes occurs during speech perception: the integration of an auditory signal with the talker's corresponding visual mouth movements. The McGurk effect (McGurk & Macdonald, 1976) highlights the importance of how these two modalities play a significant role in our perception of language. This phenomenon refers to how visual mouth information can wildly change our percept of an incoming auditory signal. When presented with an auditory stimulus of /ba/ paired with an incongruent mouth movement of the speaker producing /ga/, many participants report perceiving a fused sound such as "da" (McGurk & MacDonald, 1976). Similarly, the visual dominance illusion occurs when individuals perceive the visually conveyed phoneme (viseme) when presented with an incongruent audiovisual (AV) pairing (Abbott & Shahin, 2018; Rosenblum & Saldaña, 1992). For example, when presented with visual-/fa/ paired with auditory-/ba/, many listeners report hearing the viseme, 'fa'. This illustrates that speech perception is more than just what we hear; it is also affected by what we see.

It is important to note that while the McGurk effect is robust and occurs in many individuals, its occurrence is not universal and is influenced by individual differences (Basu Mallick et al., 2015). Prior research has found that various individual differences, such as genetics, having a natural inclination towards hearing the illusion, and lipreading abilities, play a large role in experiencing the McGurk effect (Basu Mallick et al., 2015; Brown et al., 2018; Feng et al., 2019). A study investigating the McGurk effect in professional musicians found that they were less susceptible to the illusion, potentially due to their finer acoustic processing and enhanced neural representations of speech resulting from musical training. (Proverbio et al., 2016). Additionally, children under the age of 9 years are less likely to experience the illusion than children between 10-12 years and adults between 20-35 years (Hirst et al., 2018). Indeed, the original McGurk illusion study found that children under the age of 5 years are less likely to experience the illusion compared to children between 7-8 years and adults (McGurk & Macdonald, 1976). Furthermore, it has also been discovered through eye-tracking, that individuals who naturally gaze at the mouth of a speaker more often will also experience the McGurk effect more (Gurler et al., 2015).

We strongly rely upon visual information in degraded and noisy environments to aid in clarifying incoming speech signals (Grant & Seitz, 2000; Sumby & Pollack, 1954). The presence of visual lip movements enhances speech comprehension performance compared to auditory-only conditions, when listening to speech in noise (Sumby & Pollack, 1954). When presented with degraded target speech, visual lip-reading can help fill in the gaps (Grant & Seitz, 2000). Knowing that visual information is crucial to our perception and comprehension of language and that we heavily rely upon it when

auditory input is unclear, the question arises how one's language experience (e.g., bilingualism) may interact with the use of visual information during speech perception.

In quiet, auditory-only conditions, bilinguals have been shown to be influenced by their knowledge of two languages in various phonetic perception tasks, as they use different speech cues for each language. A phonetic cue, such as voice onset time (VOT), that can enable the discrimination between phonemes in one language may not be necessarily useful in another language (Abramson & Lisker, 1973; Polka et al., 2001). Research suggests that bilinguals may have a double phonemic representation: two phonemic representations, one for each language, for a single acoustic-event (Garcia-Sierra et al., 2009). VOT is the most common phonetic cue that is implemented when testing bilinguals' double phonemic representation. This is due to its acoustic properties, allowing listeners to contrast between certain phonemes (i.e., word-initial voiced vs voiceless stop consonants) (Abramson & Lisker, 1973; Lisker & Abramson, 1970). It is interesting to note that different variations of VOT can evoke different perceptions of phonemes (Best et al., 2001; Kuhl et al., 1992; Lisker & Abramson, 1970). For example, short durations of VOT are perceived as voiced sounds by English monolinguals but as voiceless sounds by Spanish monolinguals (Lisker & Abramson, 1970). To test this double phonemic effect, many studies have been conducted, but they have yielded mixed results. Some studies have found no significant differences in bilinguals' perceptual boundaries across language contexts (Caramazza et al., 1973; Williams, 1977), while others have found evidence in favor of the double phonemic effect (Casillas & Simonet, 2018; Elman et al., 1977; Garcia-Sierra et al., 2009). Studies in favor of the effect have also found evidence of perceptual boundary shifts in bilinguals when language context is established before or throughout the study (Casillas & Simonet, 2018; Elman et al., 1977; Garcia-Sierra et al., 2009; García-Sierra et al., 2012). These results indicate that bilinguals can employ language-specific perceptual boundaries when processing speech sounds, suggesting a separation between their representation of phonemes.

However, despite this prior work, it is still not well understood how bilinguals perceive and represent phonemes across their known languages. While bilinguals can understand speech in their second language (L2) just as well as native speakers of that language in quiet conditions, they perform less accurately at speech recognition when listening to their L2 (or less dominant language) in the presence of background noise (Borghini & Hazan, 2018; Florentine, 1985; M. L. G. Lecumberri et al., 2010; Mayo et al., 1997; Takata & Nábělek, 1990). Additionally, it has been found that individuals who grew up speaking Spanish at home gain an advantage in recognizing words in noisy environments through predictive processing, whereas people who learned Spanish later in life do not gain this benefit (Fricke & Zirnstein, 2022). A potential explanation for this perceptual difference in noise could be due to cross linguistic interference—the concept that the two languages a bilingual knows are both activated when listening to speech, and words from the non-target language can hinder lexical processing of the target language (Marian & Spivey, 2003; Thierry & Wu, 2007). This cross-language interference effect can also be witnessed at the phonemic level, such that bilinguals are biased by the perceptual boundaries of their native language when categorizing non-native phonemes (Best, 1994; Ingvalson et al., 2012; Kuhl et al., 1992).

Due to less familiarity with their L2, cross linguistic interference, and having a more difficult time discerning their L2/less dominant language in adverse conditions, bilinguals are expected to rely more heavily upon visual cues than their monolingual counterparts (Chen & Hazan, 2007; Marian et al., 2018; Sekiyama, 1994). In fact, infants raised in a bilingual environment are more likely to begin to naturally gaze at the mouth of speakers than monolingual infants (Pons et al., 2015). Additionally, it has been found that Korean-English bilinguals are much more likely to experience the classic McGurk illusion than English monolinguals, again due to their tendency to gaze at the mouth more often (Marian et al., 2018). Furthermore, the presence of visual mouth movements can greatly enhance a bilingual's speech perception of their L2 (Chauvin et al., 2024; Chauvin & Phillips, 2022; Navarra & Soto-Faraco, 2007). When Spanish-Catalan bilinguals (L1 Spanish) were presented with the open-vowel /ɛ/ and the closed-vowel /e/ (a contrast in Catalan) without visual aid, the participants struggled in differentiating these vowels, as only /e/ exists in Spanish. However, once lip movements were paired with the vowels, the bilinguals were much more accurate in perceiving the two vowels (Navarra & Soto-Faraco, 2007).

While previous research has advanced our understanding of bilingual phonetic perception and the role of visual information, little research has explored how heritage bilinguals mentally represent and perceive phonemes that exist in one language but are mapped differently in another language. A heritage speaker is someone who has learned a minority language at home while growing up in a country/region with a different dominant language (Montrul, 2011; Valdés & Anderson, 2000). Prior research has looked at how bilinguals perceive phonemes differently across languages, such as the case of /r/ and /l/ in Japanese-English bilinguals. Native Japanese speakers have difficulty perceptually distinguishing between the phonemes, /r/ and /l/, whereas a native English individual is aware that these are two distinct phonemes in their native language. While research has been done looking at Japanese-English bilinguals and their categorization of /r/ and /l/, this work has mostly used bilingual populations who acquired English as a second language, and not as heritage speakers. (Ingvalson et al., 2012; MacKain et al., 1981; Miyawaki et al., 1975; Mochizuki, 1981). Of key interest for the present study are heritage speakers of Spanish (HSS) who grew up in the United States: bilinguals who grew up speaking Spanish in the home while being raised in America, where the dominant language in society is English. It is important to note that HSS are not a homogenous population; they vary greatly in terms of when and how they acquired Spanish and English (Montrul, 2011). Some HSS are children of immigrant parents who are monolingual-Spanish speakers, who may have been exposed to English only with formal schooling and/or acted as language brokers for their parents. Some HSS were exposed to both Spanish and English simultaneously, and others may have migrated to an English-dominant country at a later age in childhood. Some HSS are fairly competent in understanding and speaking Spanish, while others may barely be able to speak or understand the heritage language.

This is also an excellent population to study due to differences between Spanish and English phonology. A study done by Garcia and Cooke (2006) found that dominant Spanish-English bilinguals (first language (L1): Spanish, also their dominant language) often misidentified the English phoneme /v/ as /b/. This misclassification is interesting to

note because Spanish phonology, unlike English, does not have a distinction between the phonemes /b/ and /v/. In Spanish, the phoneme /v/ maps onto /b/, which is why often, many native Spanish speakers replace /v/ sounds in English words with /b/ (e.g., *volleyball* is pronounced like *bolleyball*). This difference in the two languages' phonologies poses an interesting question: How do Spanish-English bilinguals perceive the phonemes /b/ and /v/, which are represented differently across the two languages? Knowing the influence of visual information in bilinguals and how it can enhance L2 perception, how does the presence of mouth movements impact the precision of these representations of /b/ and /v/?

In the current study, we aim to examine how one's language experience influences their phonetic perception, under auditory-only (A-Only) and audiovisual (AV) conditions. Specifically, the study examines phonetic perception of an English-specific phoneme compared to a phoneme that is shared across both languages, to determine how phonetic context interacts with language background. Participants (n = 43) fell along a continuum of language dominance, ranging from English dominant to Spanish dominant, to understand the effects of language dominance on phonetic perception. We used a series of audiovisual stimuli that consisted of auditory stimuli along a /va/-/ba/ continuum paired with visual-/va/ or visual-/ba/. By using this continuum, we were able to measure the proportion of trials that participants perceive /ba/ or /va/ for each stimulus and examine the extent of visual influence on the auditory stimuli. Participants were presented with AV and A-only stimuli and asked to transcribe what they heard. The experimental design resulted in AV congruent trials (e.g., auditory-va paired with visual-va), AV incongruent trials (e.g., auditory-va paired with visual-ba), as well as other trials involving an ambiguous auditory stimulus paired with either visual /ba/ or visual /va/. Our research questions were: 1) Because of how Spanish phonology /v/ is represented in Spanish, under auditory-only conditions, are individuals who are less English dominant, more biased towards hearing /ba/ than English monolinguals, especially when the auditory stimulus is ambiguous?, 2) Under AV conditions, do the specific auditory and visual stimulus pairings modulate auditory perception in a language-dependent manner?

If one's early language experience can impact auditory-only phonetic processing in young adulthood, as English dominance decreases, we expect to observe more 'ba' responses, especially when the auditory stimulus is ambiguous. In terms of audiovisual perception, visual context should influence auditory perception regardless of language experience. However, this effect of visual context is expected to be more pronounced for bilingual individuals who have more balanced English-Spanish dominance. As noted above, some AV pairings are congruent, for example, when visual-/va/ is paired with auditory-/va/, while other AV pairings are incongruent (e.g., visual-/ba/ paired with auditory-/va/). Incongruent AV pairings like these, in which the viseme has a very visible place of articulation (i.e., bilabial for /b/, labiodental for /v/), can lead to the visual dominance illusion in which participants report hearing the visually-conveyed phoneme (e.g., visual-/ba/ + auditory-/va/ leads to illusory perception of /ba/) (Rosenblum & Saldaña, 1992; Shahin et al., 2018). For the incongruent pairings, we expect participants' auditory perception to be consistent with their language experience. For example, if language experience impacts speech perception, then as English dominance decreases, participants should be more likely to perceive illusory-'ba' when auditory-/va/ is paired

with visual-/ba/, since the /v/ phoneme does not exist in Spanish. Similarly, for ambiguous auditory stimuli, individuals leaning more towards balanced English-Spanish dominance on the language dominance continuum should be more likely to perceive these ambiguous sounds as 'ba' when paired with visual-/ba/ compared to those leaning more towards English dominance. Individuals with greater English dominance should be more inclined to perceive these ambiguous auditory stimuli paired with visual /va/ as 'va' compared to those with lower English dominance.

2. Materials and Methods

2.1. Participants

Fifty-seven healthy young adults participated in this study. However, fourteen participants were excluded due to technical issues during data collection, not following task instructions, not fitting the language criteria (knowing languages other than English and/or Spanish), or failure to complete the study, resulting in usable data from forty-three participants (>18 years of age, M = 20.51 years, SD = 2.41 years, 6 participants did not provide their specific age; 32 females, 9 males, 2 did not respond). The participant sample comprised both English monolinguals, as well as Heritage Spanish speakers who varied in their degree of English dominance. As described in more detail below in the data analyses section, we quantified participants' language dominance along a continuum using the Bilingual Language Profile (BLP) (Birdsong et al., 2012). All participants self-reported normal hearing, normal or corrected vision, and no language deficits. Participants were recruited via an internal recruiting system of the University of California, Merced and provided written consent prior to participation. Participants were compensated with course credit or cash. All experimental protocols were approved by the Institutional Review Board (IRB) of the University of California, Merced, and all methods were carried out in accordance with the guidelines and regulations of the IRB of the University of California, Merced and in accordance with the Declaration of Helsinki.

2.2. Stimuli

The study consisted of two main types of stimuli: auditory-only (A-only) and audiovisual (AV) stimuli. For both conditions, consonant-vowel (CV) syllables along a /va/ to /ba/-like continuum were used. This continuum was created by taking two exemplars of a female talker uttering the CV of /va/ (mean f0 = 199 Hz) and manipulating the fricative /v/ segment of the CV. The phoneme /v/ is classified as a fricative, due to friction of airflow through a narrow opening of the vocal articulators when it is produced. This results in a burst of high frequency lasting approximately 115 ms at the beginning of the /va/ tokens we used. If this fricative portion is removed, the CV is often perceived as 'ba.' In Adobe Audition (Adobe, Inc., San Jose, CA), we manipulated the /v/ fricative segment by reducing its amplitude in steps of -9 dB for a total of four times (i.e., Step 1 = original /va/ recording, Step 2 = /v/ adjusted by -9 dB, Step 3 = -18 dB, Step 4 = -27 dB, Step 5 = -36 dB). This resulted in 5 stimuli that ranged from /va/ to /ba/-like. We refer to the final stimulus as /ba/-like since it is not an actual recording of /ba/, but rather a manipulation of /va/ that sounds like /ba/. We also created a second /va/ to /ba/-like 5-step continuum for each /va/ exemplar, by reducing the /v/

fricative's duration. However, for the purposes of the present report, the analyses included herein are restricted to the trials involving the amplitude manipulation of the /v/ fricative. Each exemplar was counterbalanced across participants, to ensure that results were not driven by the way the talker produced the CV.

To create the AV stimuli, the auditory stimuli were paired with a video of the female talker producing /va/ or /ba/, using custom MATLAB code (The MathWorks, Inc., Natick, MA). There were two video exemplars for each CV. The videos were cropped, so that only the nose to the neck were visible. This was to ensure that participants would focus on the mouth movements, and not be distracted by any other physical features (e.g., her eyes). To create the AV pairings, the auditory portion of the original video was removed and replaced with one of the auditory stimuli created along the continuum. The new auditory component was temporally aligned with the original acoustic onset of that video. For the A-only stimuli, a single frame from each video, before the talker spoke, was paired with each auditory stimulus to control for any visual effects. This ensured that the frame selected had the talker's mouth in a neutral position. In total, there were 20 AV stimuli and 20 A-only stimuli (5 auditory stimulus steps along the continuum x 2 visemes x 2 video exemplars). All auditory stimuli were normalized in Adobe Audition to the same sound intensity and were presented at ~62 dBA sound pressure level.

2.3. Procedure

Participants sat in a sound attenuated booth (IAC Acoustics, Naperville, IL) about 127 cm from a 22-inch computer monitor with two external Yamaha HS8 speakers on either side, located at ±45˚ relative to the listener. A-only and AV stimuli were presented in separate blocks (5 blocks per condition, 10 blocks total). AV and A-only Block order was counterbalanced across participants (e.g., all A-only blocks were presented first followed by all AV blocks or vice versa). Each stimulus was presented twice in each block, for a total of 200 trials per block. However, as mentioned earlier, we also included another /va/ to /ba/-like 5-step continuum, by reducing the /v/ fricative's duration. With the addition of this manipulation, participants were presented with a total of 400 trials/block. Further broken down, for A-Only blocks and focusing on the amplitude manipulation, there were 40 trials for each step across all 5 blocks. For AV blocks and amplitude-manipulated stimuli, there 20 trials for each step and each viseme (/ba/ or /va/) (e.g., 20 trials for Step 1 paired with visual-/ba/ and 20 trials for Step 1 paired with visual-/va/). Prior to the start of each block, participants were told they would be presented with speech sounds, not words, that were either accompanied with or without a video (depending on the block condition). It was emphasized that participants should always be paying attention to the screen to ensure that they were focused on the talker's mouth movements. The participants' task was to type out what they heard. Participants were offered an optional break between each block.

2.4. Data and Statistical Analysis

Logfiles of participants' responses were transferred to Excel spreadsheets, which were then parsed using custom MATLAB code. Participants' responses were categorized according to the first letter transcribed by the participant (i.e., responses "ba", "bah", and "bo" were all included in the response category /b/). A table containing information about

the block condition (AV or A-only), viseme (/ba/ or /va/ video), step (which auditory stimulus along the continuum), and the first letter of the response for each trial was created for every block for each participant. Additionally, each participant's BLP score was computed from their responses on the BLP questionnaire. While the original BLP scale ranges from -218 to +218, we normalized this scale for simplicity such that BLP scores would range from -1 to 1, with -1 reflecting English dominance, 0 indexing balanced English-Spanish dominance, and 1 for Spanish dominance.

For statistical analysis, we created a binomial distribution of the response data, such that the target response was coded as 1 and all other responses were coded as 0. Since we were specifically examining speech perception along a /va/ to /ba/-like continuum, we conducted two sets of analyses—one in which 'v' was the target response and the second in which 'b' was the target response. Using R Studio (Version 4.3.2, R Core Team, 2023) and the lme4 package (Bates et al., 2015), we fitted Generalized Linear Mixed-Effects Models (GLMMs) to the data using logistic regression, since we coded the response data in a binomial manner. Post-hoc pairwise comparisons were done using the emmeans package in R (Lenth, 2024).

First, we conducted a mixed effects binomial linear regression on the A-only data to examine whether language dominance (as quantified by subjects' BLP score) and Step (the 5 auditory stimuli along the continuum) predict subjects' auditory perception. This model addresses the first research question about how language dominance affects auditory phonetic perception in a language-dependent manner. Since we were interested in perception of 'ba' or 'va' and responses were coded in a binary fashion, we ran two models, one to predict each percept. We also ran one model in which Step was a continuous variable and one where it was a categorical variable. This was done to determine whether there would be an overall effect of language dominance (when Step is continuous) and to elucidate which steps were significantly modulated by language dominance (when Step is categorical).

For 'ba' responses, this model included the fixed effects of BLP Score and Step (reference level: Step 1 (the original /va/ auditory stimulus) when categorical), as well as the intercept corresponding to each subject as a random effect. Using the function glmer, the formula was 'ba' Response ~ Step * BLP Score + 1|SubjectID. For 'va' responses, the model was the same, except the prediction outcome was changed ('va' Response ~ Step * BLP Score + 1|SubjectID). We also utilized the bobyqa optimizer. A positive effect of BLP Score would reveal that lower English dominance influences perception of the target phoneme. An effect of Step would indicate that the continuum drives perception of 'ba' or 'va.' An interaction between BLP Score and Step would suggest that both language dominance and the position along the continuum play a role in shaping perception of the target phoneme.

To address the second research question about how language dominance interacts with visual context during AV speech perception, a second set of statistical analyses was done. For these analyses, the parsed data included an additional set of coded data where a value of 1 was assigned to responses of 'ba,' 'ma,' or 'pa,' since the phonemes /b/, /m/, and /p/ share the same viseme. Similarly, responses of 'va' and 'fa' were coded as 1 because the phonemes /v/ and /f/ share the same viseme. A mixed effects binomial linear

regression was done to examine whether BLP Score, Step (the 5 auditory stimuli along the continuum), and AV condition predict subjects' auditory perception. As with the A-Only models, we initially ran a model in which Step was a numerical variable, followed by an additional model where Step is a categorical variable. For 'ba' viseme responses, this model included the fixed effects of BLP Score, Step, and AV Condition (AV-/ba/ and A-Only), as well as the intercept corresponding to each subject as a random effect. AV-/va/ was not included in this model because the aim of the study was to uncover how a specific visual mouth movements influence perception of the respective viseme. In that case, when predicting 'ba' viseme responses, it is unnecessary to include the factor of AV-va. Again, using the glmer function, the formula was 'ba' Viseme Response ~ Step * BLP Score * AV Condition + 1|SubjectID. A similar model was created for 'va' viseme response, the only difference being the prediction outcome and that the AV Condition consisted of A-Only and AV-/va/ instead of AV-/ba/ ('va' Viseme Response ~ Step * BLP Score * AV Condition + 1|SubjectID). These models would illustrate whether there are effects of language dominance, the auditory continuum, and/or AV condition on perception of 'ba/ma/pa' or 'va/fa'.

3. Results

3.1. A-Only

The primary binomial linear regression analysis was performed to examine how language dominance and the /va/-/ba/ continuum affects perception of the auditory-only stimuli. This was done by first inputting BLP Score and Step (as a numerical variable) as fixed effects into the model. This was then followed by inputting BLP Score and Step (as a categorical variable) as fixed effects into the model to further explore the interaction of the 5 different steps with BLP score. When Step was inputted as a categorical variable, Step 1 on the continuum (/va/) was set as the referent level for the outcome measure. Average frequency responses for 'ba' and 'va' across Step are plotted in Figure 1.

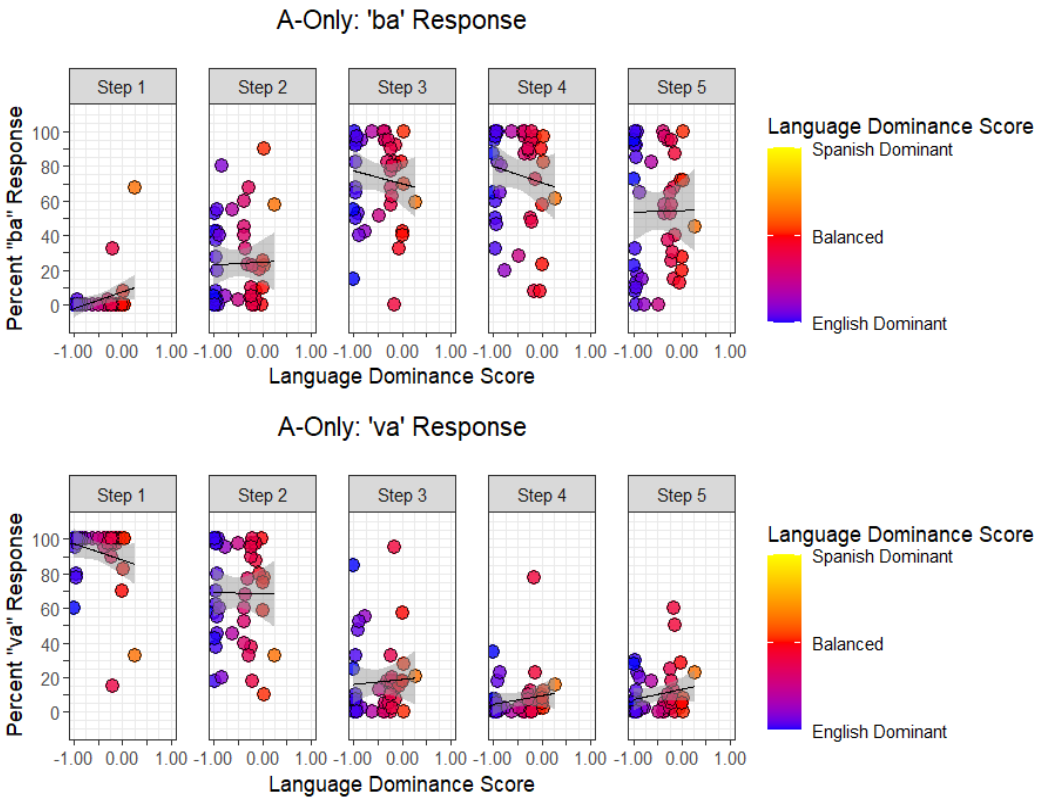

Figure 1: Average frequency responses for 'ba' (top) and 'va' (bottom) across Step as a function of Language Dominace (BLP) scores.

Starting with the model predicting ‘ba’ response when Step was treated as a numerical value, we found that there a significant main effect of Step (B = 0.77, SE = 0.03, p = <0.001) but not for BLP score (see Table 1). However, there was a significant two-way interaction between the two factors (B = -0.13, SE = 0.05, p = 0.02), suggesting that the effect of BLP score on perceiving ‘ba’ varied depending on the level of Step.

Table 1: Predicting 'ba' responses from BLP Score and Step (numeric)

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	-2.61	0.27	**<0.001*** **	-3.14	-2.08
BLP Score	0.29	0.42	0.5	-0.54	1.11
Step	0.77	0.03	**<0.001*** **	0.7	0.83
BLP Score:Step	-0.13	0.05	**0.02***	-0.23	-0.02

To understand how Step was interacting with BLP score, we followed up with a similar model, but now Step was treated as categorical variable (see Table 2). For predicting ‘ba’ response, there was a significant main effect of BLP score with a positive estimate value, suggesting that a higher BLP score predicts greater perception of ‘ba’ responses at Step 1 (the reference level of Step). Additionally, there were significant main effects for all Steps relative to Step 1, all with positive estimates. Overall, as the auditory stimulus becomes more ‘ba’-like, participants were more likely to perceive ‘ba’ than the original /va/ stimulus (Step 1). There were also significant two-way interactions between each Step and BLP Score relative to Step 1 (/va/ stimulus), which can be seen in the slopes plotted for each Step in Figure 2. Compared to the results of Step 1, the slopes of ‘ba’ response for each Step across BLP score is different, with some slopes being steeper and/or negative. These results were then followed-up by a post-hoc pairwise comparisons test of the slopes using the emtrends function from the emmeans library in R (see Table 3). Results indicated only contrasts between Step 1 and all other Steps were significant. This is further validated in the interaction plot (Figure 2), in which only the slope for Step 1 across BLP score has the most dramatic slope. This indicates that the interaction between BLP Score and Step for ‘ba’ response was being driven by Step 1. Thus, as BLP score increases (i.e., English dominance decreases), participants were significantly more likely to report hearing the Step 1 /va/ stimulus as ‘ba’, but this BLP effect was not observed at any of the other Steps.

Table 2: Predicting A-Only 'ba' responses from BLP Score and Step (categorical)

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	-2.99	0.36	<**0.001*****	-3.69	-2.29
BLP Score	8.4	1.09	<**0.001*****	6.27	10.53
Step 2	1.5	0.21	<**0.001*****	1.09	1.92
Step 3	4.07	0.22	<**0.001*****	3.64	4.5
Step 4	4.12	0.22	<**0.001*****	3.69	4.55
Step 5	3.16	0.21	<**0.001*****	2.74	3.57
BLP Score:Step 2	-8.4	0.99	<**0.001*****	-10.34	-6.45
BLP Score:Step 3	-8.83	0.99	<**0.001*****	-10.77	-6.89
BLP Score:Step 4	-8.97	0.99	<**0.001*****	-10.92	-7.03
BLP Score:Step 5	-8.41	0.99	<**0.001*****	-10.35	-6.48

Table 3: Estimated marginal means from the A-Only 'ba' model of the interaction between Step and BLP Score (significant results only)

Contrast	Estimate	SE	z.ratio	p-value
Step 1 - Step 2	8.4	0.99	8.46	**<0.001*** **
Step 1 - Step 3	8.83	0.99	8.90	**<0.001*** **
Step 1 - Step 4	8.97	0.99	9.04	**<0.001*** **
Step 1 - Step 5	8.41	0.99	8.51	**<0.001*** **

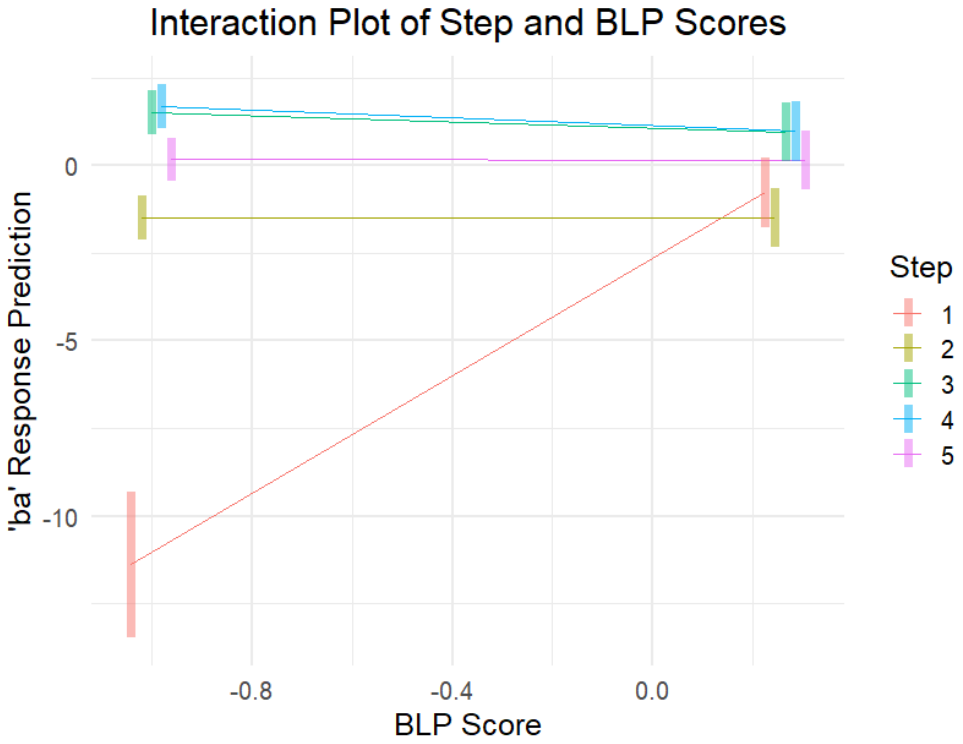

Figure 2: Estimated trends from the A-Only 'ba' perception model depicting the interaction between BLP Score and Step

We ran the same model for predicting 'va' perception, where Step was treated as a numerical (see Table 4). The model revealed significant main effects for both BLP Score (B = -1.67, SE = 0.53, p = 0.002) and Step (B = -1.44, SE = 0.05, p = <0.001; See Table 2). However, there was also a significant two-way interaction between the two factors (B = 0.69, SE = 0.1, p = <0.001). Similar to the results from the 'ba' model, this suggests that the effect of BLP score on perceiving 'va' varied depending on the level of Step.

Table 4: Predicting A-Only 'va' responses from BLP Score and Step (numeric)

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	3.52	0.33	**<0.001*** **	2.86	4.17
BLP Score	-1.67	0.53	**0.002** **	-2.71	-0.62
Step	-1.44	0.05	**<0.001*** **	-1.54	-1.34
BLP Score:Step	0.69	0.1	**<0.001*** **	0.5	0.88

To understand how Step was interacting with BLP score, we followed up with a similar model, but now Step was treated as categorical variable (see Table 5). There was a significant main effect of BLP score but with a negative estimate value, suggesting that a lower BLP score (i.e., greater English dominance) predicts greater perception of 'va' responses at Step 1 (the original /va/ stimulus). Additionally, there were significant main effects at each Step, all with negative estimates relative to Step 1, suggesting that as the auditory stimulus becomes more 'ba'-like, 'va' perception decreases. This model also revealed significant Step by BLP Score interactions at Steps 2-5 relative to Step 1, which is reflected in the different slopes compared to Step 1 in Figure 1. Post-hoc pairwise analyses using the emtrends command were done to compare the 'va' response slopes as a function of BLP score between each pair of Steps (see Table 6). We observed significant differences in slope between Step 1 and all other Steps, as well as for Step 2 compared to Steps 4 and 5. This is reflected in the interaction plot (Figure 3), as Step 1 has the steepest and most negative slope compared to all other Steps across BLP score, reflecting that as BLP Score increases, participants were significantly less likely to report hearing the Step 1 stimulus as 'va'. Compared to the slope of Step 2, the slopes of Steps 4 and 5 are steeper and more positive, indicating that as BLP Score increases, at Steps 4 and 5, participants are significantly more likely to perceive 'va' than at Step 2.

Table 5: Predicting A-Only 'va' responses from BLP Score and Step (categorical)

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	2.53	0.36	**<0.001*** **	1.83	3.24
BLP Score	-1.48	0.59	**0.012***	-2.63	-0.34
Step 2	-1.55	0.17	**<0.001*** **	-1.88	-1.21
Step 3	-4.35	0.20	**<0.001*** **	-4.73	-3.96
Step 4	-5.32	0.23	**<0.001*** **	-5.77	-4.87

Step 5	-4.85	0.21	**<0.001*** **	-5.26	-4.43
BLP Score: Step 2	1.32	0.31	**<0.001*** **	0.71	1.92
BLP Score: Step 3	1.90	0.36	**<0.001*** **	1.20	2.60
BLP Score: Step 4	2.60	0.42	**<0.001*** **	1.78	3.43
BLP Score: Step 5	2.51	0.39	**<0.001*** **	1.74	3.28

Table 6: Estimated marginal means from the A-Only 'va' model of the interaction between Step and BLP Score (significant results only)

Contrast	Estimate	SE	z.ratio	p-value
Step 1 - Step 2	-1.31	0.31	-4.27	**<0.001*** **
Step 1 - Step 3	-1.90	0.36	-5.27	**<0.001*** **
Step 1 - Step 4	-2.60	0.42	-6.15	**<0.001*** **
Step 1 - Step 5	-2.1	0.39	-6.39	**<0.001*** **
Step 2 - Step 4	-1.29	0.34	-3.71	**0.0019** **
Step 2 - Step 5	-1.20	0.31	-3.86	**0.001** **

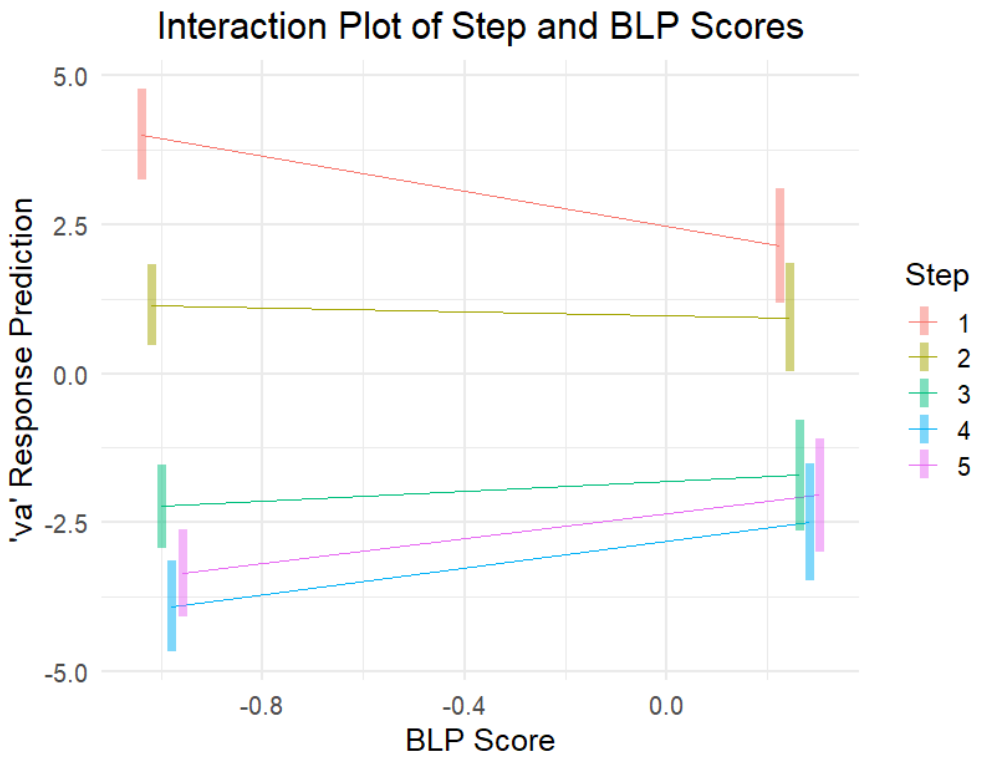

Figure 3: Estimated trends from the A-Only 'va' perception model depicting the interaction between BLP Score and Step

3.2. AV Condition

A binomial linear regression analysis was performed to examine how language dominance, the /va/-/ba/ continuum, and AV condition affect perception of 'ba' and 'va' visemes. This was done by inputting BLP Score, Step, and AV Condition (AV-ba and A-Only, or AV-va and A-Only) as fixed effects into the model.

3.2.1. 'ba/ma/pa' Model

For predicting 'ba/ma/pa' responses, remember that in the model, the fixed effect of AV Condition consisted of the levels AV-ba and A-Only. We began with the model in which Step was a numerical variable (see Table 7). The model revealed significant main effects of Step (B = 0.77, SE = 0.03, p = <0.001) and AV-ba condition (B = 0.66, SE = 0.18, p = <0.001). However, there were two significant two-way interactions between BLP Score and AV-ba condition (B = 1.06, SE = 0.31, p = <0.001) and between Step and AV-ba condition (B = 0.31, SE = 0.07, p = <0.001). Lastly, there was a three-way interaction between all three factors (B = -0.36, SE = 0.11, p = <0.001). These results suggest that the three factors interact such that the effect of perceiving 'ba/ma/pa' in the AV-ba condition is influenced by both the BLP Score and Step progression.

Table 7: Predicting AV 'ba/ma/pa' responses from BLP Score, Step (numeric), and AV condition

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	-2.46	0.24	**<0.001*** **	-2.93	-1.99
BLP Score	0.38	0.38	0.31	-0.36	1.11
Step	0.77	0.03	**<0.001*** **	0.71	0.83
AV-ba	0.66	0.18	**<0.001*** **	0.3	1.03
BLP Score:Step	-0.1	0.05	0.052 **.**	-0.2	0.0008
BLP Score:AV-ba	1.06	0.31	**<0.001*** **	0.46	1.66
Step:AV-ba	0.31	0.07	**<0.001*** **	0.17	0.44
BLP Score:Step:AV-ba	-0.36	0.11	**0.001** **	-0.58	-0.14

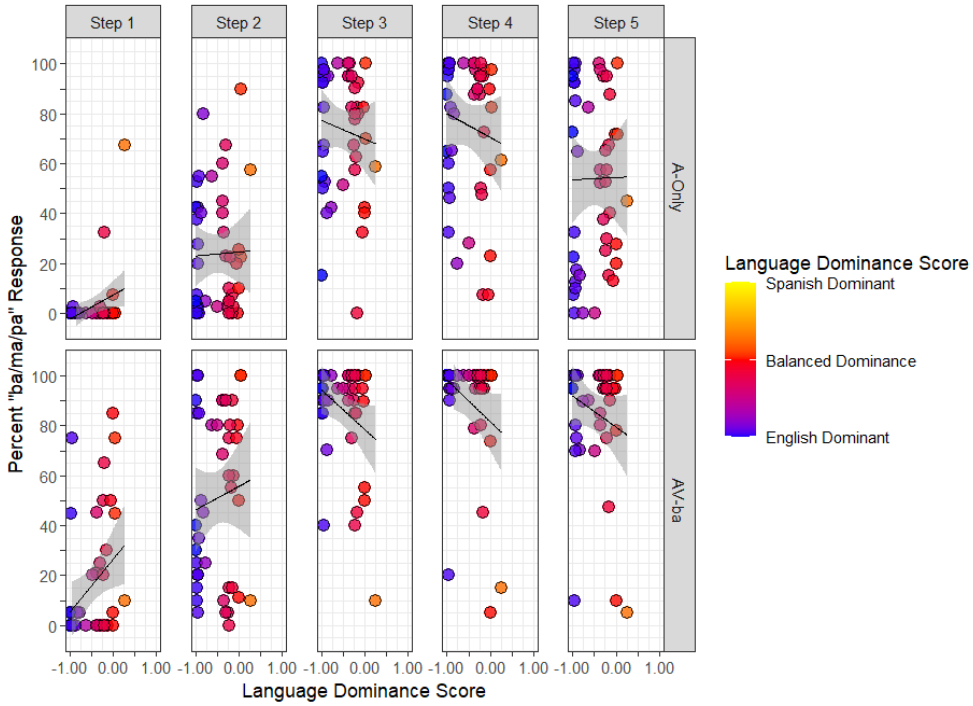

Figure 4: Average frequency of 'ba/ma/pa' response in AV-ba and A-Only conditions

To gain deeper insight into these interactions, we subsequently ran the model with Step treated as a categorical variable (see Table 8 and Figure 4). Step 1 on the continuum (original /va/) and A-Only were set as the referent levels for the Step and AV condition, respectively. For predicting 'ba/ma/pa' viseme responses, the model revealed the following significant results: main effects of BLP Score, AV-ba condition, and all Steps relative to Step 1; two-way interactions between BLP Score and AV-ba, BLP Score and Step 2, 3, 4, and 5, respectively, AV-ba and Step 3, and AV-ba and Step 5; and a three-way interaction between BLP Score, AV-ba, and all Steps.

Table 8: Predicting AV 'ba' responses from BLP Score, Step (categorical), and AV condition

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	-2.71	0.31	**<0.001*** **	-3.33	-2.10
BLP Score	9.75	1.16	**<0.001*** **	7.49	12.02
AV-ba	1.49	0.21	**<0.001*** **	1.09	1.90
Step 2	3.89	0.21	**<0.001*** **	3.48	4.29

Step 3	3.95	0.21	**<0.001*** **	3.54	4.36
Step 4	3.1	0.20	**<0.001*** **	2.71	3.50
Step 5	1.82	0.23	**<0.001*** **	1.38	2.26
BLP Score: AV-ba	-9.59	1.10	**<0.001*** **	-11.75	-7.43
BLP Score: Step 2	-10.03	1.10	**<0.001*** **	-12.18	-7.88
BLP Score: Step 3	-10.12	1.10	**<0.001*** **	-12.27	-7.96
BLP Score: Step 4	-9.53	1.10	**<0.001*** **	-11.68	-7.38
BLP Score: Step 5	-7.99	1.12	**<0.001*** **	-10.19	-5.81
AV-ba: Step 2	-0.14	0.27	0.61	-0.68	0.40
AV-ba: Step 3	-0.99	0.30	**<0.001*** **	-1.58	-0.41
AV-ba: Step 4	-0.35	0.32	0.27	-0.99	0.28
AV-ba: Step 5	0.66	0.32	**0.04***	0.03	1.30
BLP Score: AV-ba: Step 2	8.38	1.15	**<0.001*** **	6.13	10.63
BLP Score: AV-ba: Step 3	7.37	1.17	**<0.001*** **	5.08	9.67
BLP Score: AV-ba: Step 4	7.61	1.19	**<0.001*** **	5.27	9.95
BLP Score: AV-ba: Step 5	8.17	1.17	**<0.001*** **	5.87	10.47

Post-hoc pairwise contrasts were done to determine what drove the various interactions. Using emtrends, for the interaction between AV condition and BLP Score, there was a significant difference in slopes between AV-ba and A-Only (B = 1.69, SE = 0.258, p = <0.001), such that as BLP score increased (shifted towards more Spanish dominance), perception of 'ba' visemes drastically increased more in the A-Only condition than the AV-ba condition (see Figure 5). This hints at a difference in the way the phonemes /b/ and /v/ are perceived across language dominance scores. Additionally, in Figure 5, the data has been averaged across Step, therefore obfuscating detailed effects of visual information across the continuum.

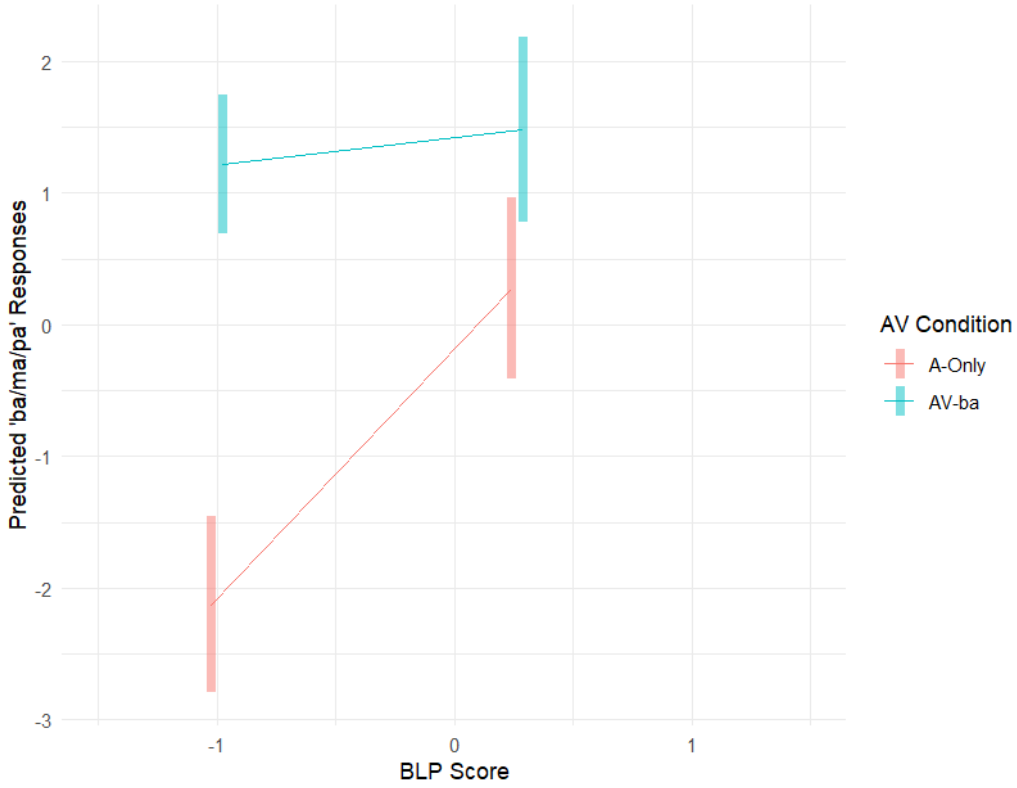

Figure 5: Estimated trends from the AV-ba model depicting the interaction between BLP Score and AV Condition.

Examining the interaction between AV condition and Step using emmeans revealed significant differences in the proportion of 'ba/ma/pa' responses between the A-only and AV-ba conditions at all five steps, with the AV-ba condition eliciting the most responses. Furthermore, contrasts between each pair of steps separately for the A-Only and AV-ba conditions unveiled significant differences in 'ba/ma/pa' responses at each pair of steps except for Step 3 and 4 in the A-Only and AV-ba conditions, Step 4 and 5 in the AV-ba condition, and Step 4 and 5 in the AV-ba condition (see Table 9). These contrasts are visualized in Figure 6, which shows the estimated marginal means for each Step and AV condition combination. As the auditory stimulus became more 'ba'-like, perception of 'ba' increased for both A-Only and AV-ba conditions, especially so at Step 1. This was due to Step 1 being the /va/ auditory stimulus, where in A-Only conditions, would mostly be perceived as 'va'; as the auditory stimulus moved toward the right of the continuum, perception of 'ba' would increase. With the addition of AV-ba, this increase in 'ba' response was strongest at Step 1, compared to all other Steps between both AV conditions, indicating that at times, participants experienced the visual dominance illusion (auditory /va/ + visual-ba = perceiving 'ba/ma/pa').

Table 9: Estimated marginal means from the AV-ba model of the interaction between Step and AV Condition

Step	AV Condition	Contrast	Estimate	SE	z.ratio	p-value
1	.	AV-ba - A-Only	5.85	0.62	9.43	**<0.001*** **
2	.	AV-ba - A-Only	1.48	0.10	15.24	**<0.001*** **
3	.	AV-ba - A-Only	1.14	0.13	8.49	**<0.001*** **
4	.	AV-ba - A-Only	1.66	0.16	10.35	**<0.001*** **
5	.	AV-ba - A-Only	2.39	0.14	17.25	**<0.001*** **
.	A-Only	Step 2 - Step 1	6.33	0.62	10.29	**<0.001*** **
.	A-Only	Step 3 - Step 1	8.94	0.62	14.50	**<0.001*** **
.	A-Only	Step 3 - Step 2	2.62	0.09	29.63	**<0.001*** **
.	A-Only	Step 4 - Step 1	9.05	0.62	14.67	**<0.001*** **
.	A-Only	Step 4 - Step 2	2.72	0.09	30.41	**<0.001*** **
.	A-Only	Step 4 - Step 3	0.11	0.09	1.21	1.00
.	A-Only	Step 5 - Step 1	7.91	0.62	12.84	**<0.001*** **
.	A-Only	Step 5 - Step 2	1.58	0.08	19.38	**<0.001*** **
.	A-Only	Step 5 - Step 3	-1.04	0.08	-12.67	**<0.001*** **
.	A-Only	Step 5 - Step 4	-1.14	0.08	-13.77	**<0.001*** **
.	AV-ba	Step 2 - Step 1	1.96	0.13	15.52	**<0.001*** **
.	AV-ba	Step 3 - Step 1	4.23	0.16	26.68	**<0.001*** **
.	AV-ba	Step 3 - Step 2	2.27	0.14	15.95	**<0.001*** **
.	AV-ba	Step 4 - Step 1	4.86	0.18	26.80	**<0.001*** **
.	AV-ba	Step 4 - Step 2	2.90	0.17	17.34	**<0.001*** **
.	AV-ba	Step 4 - Step 3	0.63	0.19	3.32	**0.02***
.	AV-ba	Step 5 - Step 1	4.45	0.16	27.04	**<0.001*** **

.	AV-ba	Step 5 - Step 2	2.49	0.15	16.72	**<0.001*** **
.	AV-ba	Step 5 - Step 3	0.22	0.17	1.25	1.00
.	AV-ba	Step 5 - Step 4	-0.41	0.19	-2.12	0.86

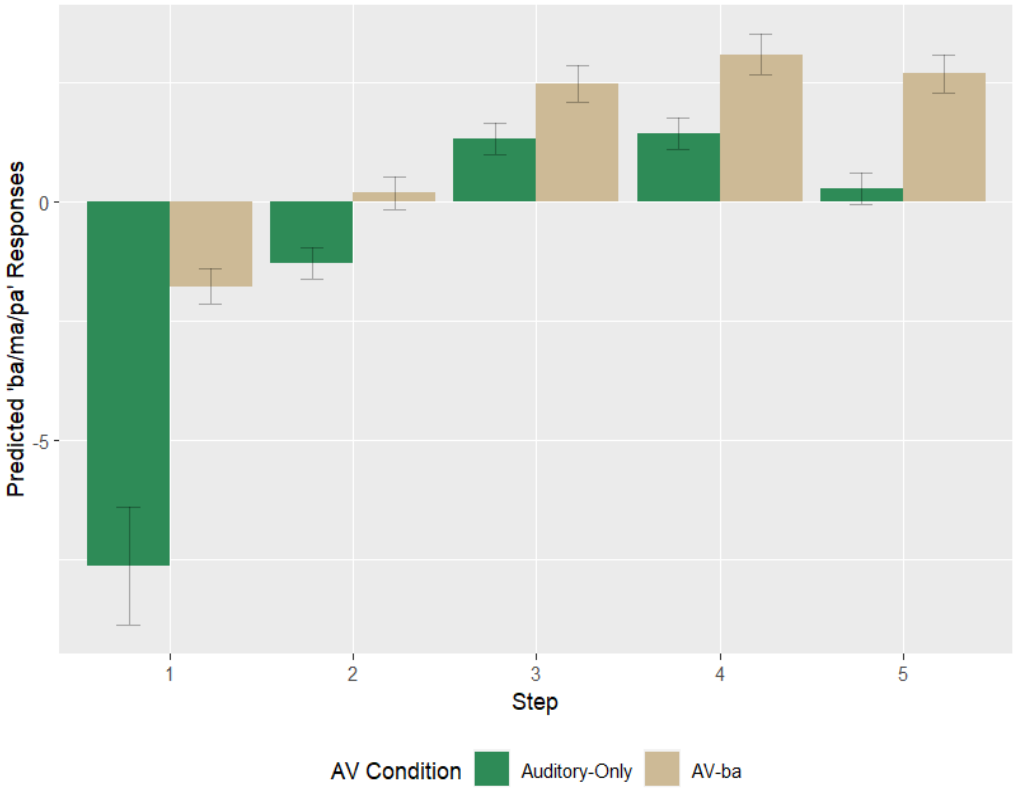

Figure 6: Plot of estimated marginal means from the AV-ba model depicting the interaction between Step and AV Condition

Pairwise contrasts (using emtrends) to decompose the interaction between BLP Score and Step showed significant differences in slopes for all pairings except Step 2 vs Step 5 and Step 3 vs Step 4 (see Table 10). Figure 7 illustrates these relationships, demonstrating parallel slopes between Step 2 and Step 5, as well as between Step 3 and Step 4, indicating that between Steps 2 and 5, and Steps 3 and 4, there is so significant difference in 'ba' response across language dominance. Figure 7 also depicts Step 1 having the steepest and most positive slope compared to all other Steps across BLP score, reflecting that as BLP Score increases, participants were significantly more likely to report hearing the Step 1 stimulus as 'ba'.

Table 10: Estimated trends from the AV-ba model of the interaction between Step and BLP Score

Contrast	Estimate	SE	z.ratio	p-value
Step 1 - Step 2	5.40	0.57	9.405	**<0.001*** **
Step 1 - Step 3	6.35	0.58	10.895	**<0.001*** **
Step 1 - Step 4	6.31	0.59	10.669	**<0.001*** **
Step 1 - Step 5	5.44	0.59	9.296	**<0.001*** **
Step 2 - Step 3	0.95	0.21	4.524	**<0.001*** **
Step 2 - Step 4	0.91	0.24	3.857	**0.001** **
Step 2 - Step 5	0.04	0.21	0.197	0.9997
Step 3 - Step 4	-0.03	0.26	-0.136	0.9999
Step 3 - Step 5	-0.90	0.24	-3.828	**0.001** **
Step 4 - Step 5	-0.87	0.26	-3.349	**0.007** **

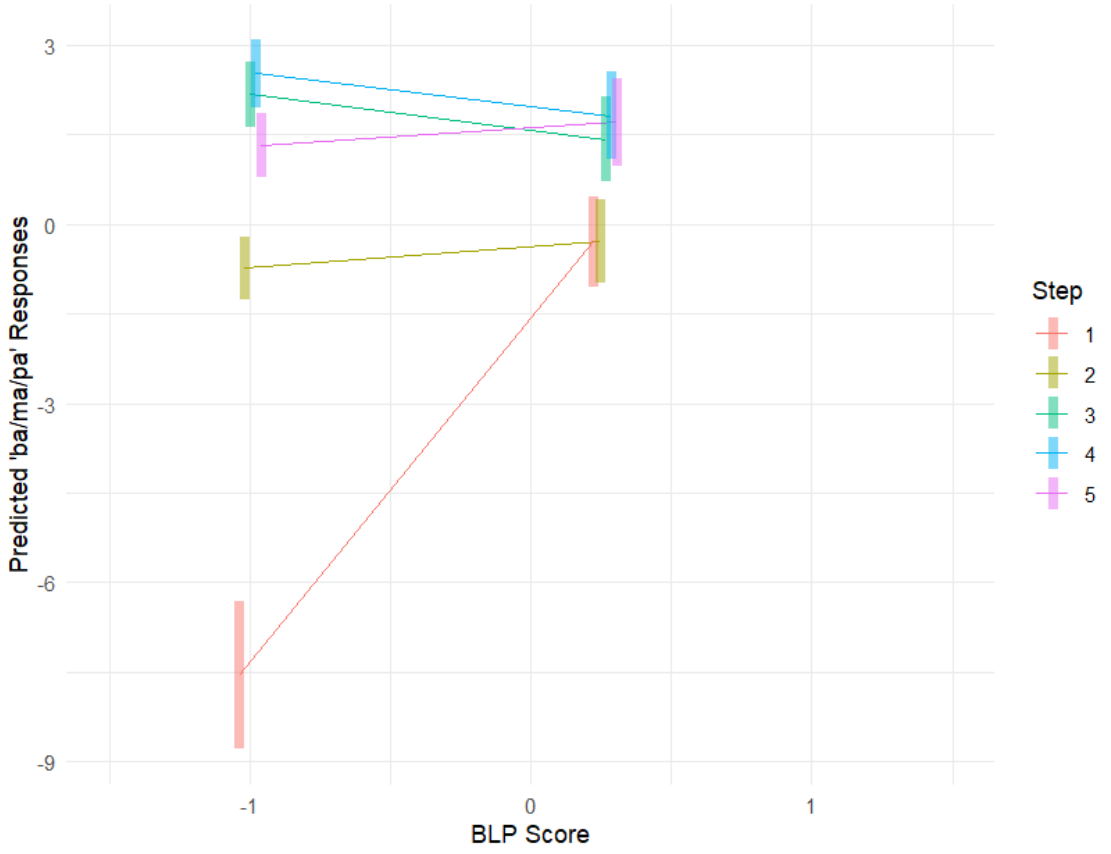

Figure 7: Plot of estimated trends from the AV-ba model depicting the interaction between Step and BLP Score

Results from the post-hoc pairwise comparisons for the three-way interaction between BLP Score, AV condition, and Step can be seen in Table 11. Referring back to Figure 4, it is evident that perceiving ‘ba/ma/pa’ is dependent upon all three factors. Differences in ‘ba/ma/pa’ responses as a function of language dominance (BLP score) and AV condition were most prominent at Step 1. This was likely since Step 1 was the pure /va/ stimulus. Perceiving ‘ba/ma/pa’ at this point along the continuum would be very unlikely. Perceiving ‘ba/ma/pa’ at Step 1 paired with visual-ba would be a case of experiencing the visual dominance illusion. As seen in Table 11, the contrast between AV-ba and A-Only at Step 1 yields a negative estimate value, indicating that as BLP score decreases, the difference in ‘ba/ma/pa’ response between AV conditions increases. However, referring to Figure 4, we observe that, on average, 'ba/ma/pa' perception between A-Only and AV-ba conditions is not significantly different at the higher end of English language dominance. As English language dominance decreases, the difference in 'ba/ma/pa' perception between AV conditions becomes much larger, with a higher perception of 'ba/ma/pa' in the AV-ba condition. It is interesting to note that at Steps 3 and 4, across AV conditions, lower English dominance results in a decrease in ‘ba’ viseme response.

Table 11: Estimated trends from the AV-ba model of the 3-way interaction between Step, BLP Score, and AV Condition

Step	AV Condition	Contrast	Estimate	SE	z.ratio	p-value
1	.	Visual-ba - A-Only	-8.00	1.12	-7.15	**<0.001*** **
2	.	Visual-ba - A-Only	0.38	0.25	1.56	1.00
3	.	Visual-ba - A-Only	-0.62	0.33	-1.91	1.00
4	.	Visual-ba - A-Only	-0.39	0.39	-0.99	1.00
5	.	Visual-ba - A-Only	0.17	0.34	0.50	1.00
.	A-Only	Step 2 - Step 1	-9.59	1.10	-8.71	**<0.001*** **
.	A-Only	Step 3 - Step 1	-10.03	1.10	-9.14	**<0.001*** **
.	A-Only	Step 3 - Step 2	-0.44	0.22	-2.00	1.00
.	A-Only	Step 4 - Step 1	-10.12	1.10	-9.20	**<0.001*** **
.	A-Only	Step 4 - Step 2	-0.53	0.22	-2.34	0.48
.	A-Only	Step 4 - Step 3	-0.08	0.22	-0.39	1.00
.	A-Only	Step 5 - Step 1	-9.53	1.10	-8.69	**<0.001*** **
.	A-Only	Step 5 - Step 2	0.06	0.21	0.31	1.00
.	A-Only	Step 5 - Step 3	0.51	0.20	2.50	0.31
.	A-Only	Step 5 - Step 4	0.59	0.21	2.87	0.10
.	Visual-ba	Step 2 - Step 1	-1.21	0.32	-3.74	**0.004** **
.	Visual-ba	Step 3 - Step 1	-2.66	0.40	-6.71	**<0.001*** **
.	Visual-ba	Step 3 - Step 2	-1.45	0.35	-4.15	**<0.001*** **
.	Visual-ba	Step 4 - Step 1	-2.51	0.45	-5.55	**<0.001*** **
.	Visual-ba	Step 4 - Step 2	-1.30	0.41	-3.16	**0.04***
.	Visual-ba	Step 4 - Step 3	0.15	0.46	0.33	1.00
.	Visual-ba	Step 5 - Step 1	-1.36	0.41	-3.27	**0.03***

.	Visual-ba	Step 5 - Step 2	-0.15	0.37	-0.40	1.00
.	Visual-ba	Step 5 - Step 3	1.30	0.43	3.05	0.06
.	Visual-ba	Step 5 - Step 4	1.15	0.48	2.41	0.40

3.2.2. 'va/fa' Model

As with the 'ba/ma/pa' model, for predicting 'va/fa' responses, we began with the model in which Step was a numerical value (see Table 12). For 'va/fa' responses, the fixed effect of AV Condition consisted of the levels AV-va and A-Only. We found significant main effects for BLP Score (B = -1.39, SE = 0.53, p = 0.008) and Step (B = -1.22, SE = 0043, p = <0.001), but not for AV condition. There were significant two and three-way interactions for all fixed effects. These results indicate that the three fixed effects interact in a way that the effect of perceiving 'va/fa' in the AV-va condition is influenced by both the BLP Score and the Step progression.

Table 12: Predicting 'va/fa' responses from BLP Score, Step (numeric), and AV condition

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	3.07	0.33	**<0.001*** **	2.42	3.73
BLP Score	-1.39	0.53	**0.008** **	-2.43	-0.36
Step	-1.22	0.04	**<0.001*** **	-1.31	-1.13
AV-va	-0.02	0.23	0.92	-0.47	0.42
BLP Score:Step	0.58	0.08	**<0.001*** **	0.42	0.73
BLP Score:AV-va	1.18	0.36	**<0.001*** **	0.48	1.88
Step:AV-va	0.79	0.07	**<0.001*** **	0.66	0.93
BLP Score:Step:AV-va	-0.45	0.11	**<0.001*** **	-0.67	-0.24

An additional model was run where Step was now a categorical variable (see Table 13 and Figure 8). For this model, the fixed effect of AV Condition consisted of the levels AV-va and A-Only, with Step 1 and A-Only set as the referent levels for the Step and AV condition, respectively. The model revealed significant results for all main effects, two-way, and three-way interactions.

Table 13: Predicting 'va/fa' responses from BLP Score, Step (categorical), and AV condition

Fixed Effect	Estimate	SE	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	2.50	0.37	**<0.001***	1.79	3.22
BLP Score	-1.61	0.60	**0.007****	-2.78	-0.44
Step 2	-1.53	0.17	**<0.001***	-1.88	-1.19
Step 3	-4.28	0.19	**<0.001***	-4.65	-3.91
Step 4	-4.96	0.21	**<0.001***	-5.37	-4.55
Step 5	-4.22	0.19	**<0.001***	-4.59	-3.85
AV-va	0.93	0.30	**0.002****	0.34	1.52
BLP Score:Step 2	1.48	0.32	**<0.001***	0.85	2.11
BLP Score:Step 3	1.97	0.35	**<0.001***	1.28	2.65
BLP Score:Step 4	2.83	0.40	**<0.001***	2.05	3.61
BLP Score:Step 5	2.35	0.35	**<0.001***	1.66	3.05
BLP Score:AV-va	1.69	0.49	**<0.001***	0.73	2.65
Step 2:AV-va	0.74	0.38	**0.05***	0.00	1.47
Step 3:AV-va	1.89	0.35	**<0.001***	1.21	2.58

Step 4:AV-va	2.63	0.36	**<0.001***	1.93	3.34
Step 5:AV-va	2.18	0.35	**<0.001***	1.49	2.87
BLP Score:Step 2:AV-va	-1.59	0.60	**0.007****	-2.76	-0.41
BLP Score:Step 3:AV-va	-2.01	0.57	**<0.001***	-3.12	-0.90
BLP Score:Step 4:AV-va	-2.75	0.59	**<0.001***	-3.92	-1.59
BLP Score:Step 5:AV-va	-1.95	0.57	**<0.001***	-3.07	-0.84

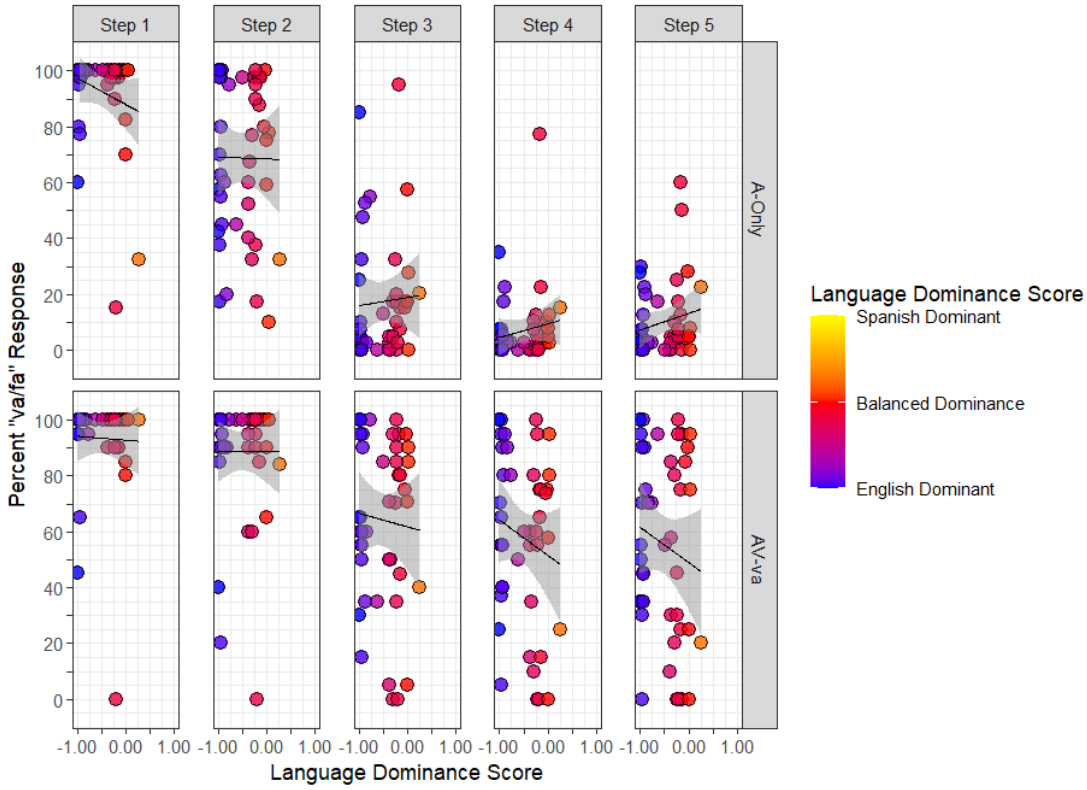

Figure 8: Average frequency of 'va/fa' response in AV-ba and A-Only conditions

Post-hoc pairwise contrasts were performed to identify the factors contributing to the observed interactions. For the interaction between AV condition and BLP Score, there

was no significant effect between AV-va and A-Only (B = -0.03, SE = 0.17, p = 0.9). While the presence of visual-/va/ boosts ‘va’ viseme perception, the strength of this effect was the same across the language dominance continuum.

Using emmeans, pairwise comparisons were conducted to explore the interaction between Step and AV Condition (see Table 14), which resulted in significant effects for all contrasts except for the AV condition effect (AV-va vs. A-Only) at Step 1, Step 5 compared to Step 3 in the A-Only condition, Step 4 compared to Step 3 in the AV-va condition, and Step 5 compared to Step 3 and 4 in the AV-va condition. These findings are visualized in Figure 9. In general, as the auditory stimulus becomes more ‘ba’-like from Step 1 to Step 5, ‘va/fa’ responses decline. The presence of visual-/va/ greatly increases the likelihood of perceiving ‘va/fa’ at Steps 2, 3, 4, and 5, but the effect plateaus at Steps 3, 4, and 5, as there are no significant differences between these three conditions in the AV-va condition. Between AV conditions, at Step 1, there is no significant difference in ‘va/fa’ response, as Step 1 is the original /va/ auditory stimulus. At Step 1 subjects are already responding ‘va/fa’ at ceiling, so the pairing of AV-va does not facilitate their auditory perception. However, as the auditory continuum moves towards ‘ba’-like, the presence of visual-va increases the proportion of 'va/fa' stimuli in the AV condition relative to the A-only condition.

Table 14: Estimated marginal means from the AV-va model of the interaction between Step and AV Condition

Step	AV Condition	Contrast	Estimate	SE	z.ratio	p-value
1	.	AV-va - A-Only	0.07	0.20	0.38	1.00
2	.	AV-va - A-Only	1.61	0.14	11.79	**<0.001*** **
3	.	AV-va - A-Only	2.98	0.11	26.23	**<0.001*** **
4	.	AV-va - A-Only	4.10	0.14	29.78	**<0.001*** **
5	.	AV-va - A-Only	3.24	0.12	27.56	**<0.001*** **
.	A-Only	Step 2 - Step 1	-2.28	0.13	-17.30	**<0.001*** **
.	A-Only	Step 3 - Step 1	-5.27	0.14	-36.77	**<0.001*** **
.	A-Only	Step 3 - Step 2	-2.99	0.10	-30.77	**<0.001*** **
.	A-Only	Step 4 - Step 1	-6.39	0.16	-39.16	**<0.001*** **
.	A-Only	Step 4 - Step 2	-4.11	0.12	-33.04	**<0.001*** **
.	A-Only	Step 4 - Step 3	-1.12	0.12	-8.97	**<0.001*** **

.	A-Only	Step 5 - Step 1	-5.41	0.15	-37.24	**<0.001*** **
.	A-Only	Step 5 - Step 2	-3.13	0.10	-31.32	**<0.001*** **
.	A-Only	Step 5 - Step 3	-0.14	0.10	-1.32	1.00
.	A-Only	Step 5 - Step 4	0.98	0.13	7.80	**<0.001*** **
.	AV-va	Step 2 - Step 1	-0.74	0.20	-3.72	0.005**
.	AV-va	Step 3 - Step 1	-2.36	0.18	-13.12	**<0.001*** **
.	AV-va	Step 3 - Step 2	-1.62	0.15	-10.87	**<0.001*** **
.	AV-va	Step 4 - Step 1	-2.36	0.18	-13.12	**<0.001*** **
.	AV-va	Step 4 - Step 2	-1.62	0.15	-10.88	**<0.001*** **
.	AV-va	Step 4 - Step 3	0.00	0.12	-0.01	1.00
.	AV-va	Step 5 - Step 1	-2.24	0.18	-12.37	**<0.001*** **
.	AV-va	Step 5 - Step 2	-1.50	0.15	-9.98	**<0.001*** **
.	AV-va	Step 5 - Step 3	0.12	0.12	1.03	1.00
.	AV-va	Step 5 - Step 4	0.12	0.12	1.04	1.00

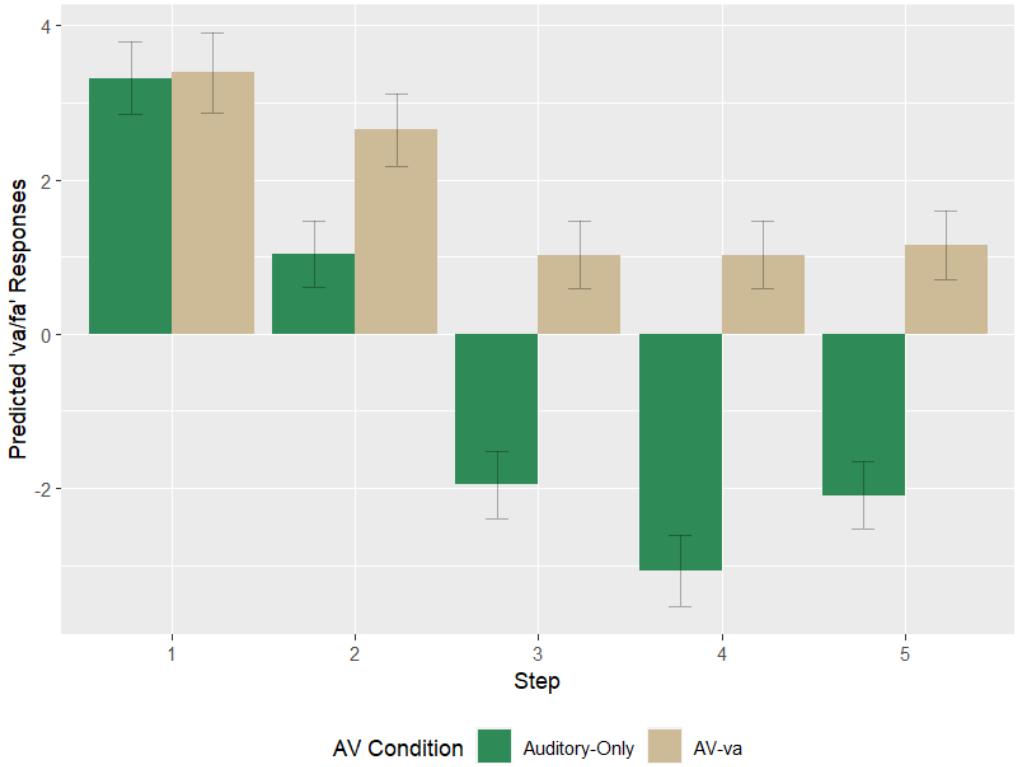

Figure 9: Plot of estimated marginal means from the AV-ba model depicting the interaction between Step and AV Condition

Using emtrends, pairwise comparisons were conducted to explore the interaction between BLP Score and Step. This revealed significant differences in slopes between Step 2 and Step 4 and 5, as well as between Step 1 and Steps 3, 4, and 5 (Table 15, Figure 10). Compared to the slopes of Steps 3, 4, and 5, the slope of Step 1 is much more negative and steeper, indicating that as English dominance decreases, participants are significantly less likely to perceive 'va/fa' at Step 1 than at Steps 3, 4, and 5. Compared to Step 2, the slopes of Step 4 and 5 are much steeper and more positive, revealing that as English dominance decreases, participants are significantly more likely to perceive 'va/fa' as Steps 4 and 5 than at Step 2.

Table 15: Estimated trends from the AV-va model of the interaction between Step and AV Condition

Contrast	Estimate	SE	z.ratio	p-value
Step 1 -Step 2	-0.6855	0.304	-2.255	0.16
Step 1 -Step 3	-0.9594	0.293	-3.272	**0.01****
Step 1 -Step 4	-1.4525	0.308	-4.718	**<0.001*****

Step 1 -Step 5	-1.3753	0.295	-4.655	**<0.001*** **
Step 2 -Step 3	-0.2739	0.226	-1.21	0.75
Step 2 -Step 4	-0.767	0.245	-3.129	**0.02** **
Step 2 -Step 5	-0.6898	0.229	-3.007	**0.02** **
Step 3 -Step 4	-0.4931	0.213	-2.312	0.14
Step 3 -Step 5	-0.4158	0.197	-2.112	0.22
Step 4 -Step 5	0.0772	0.216	0.358	1.0

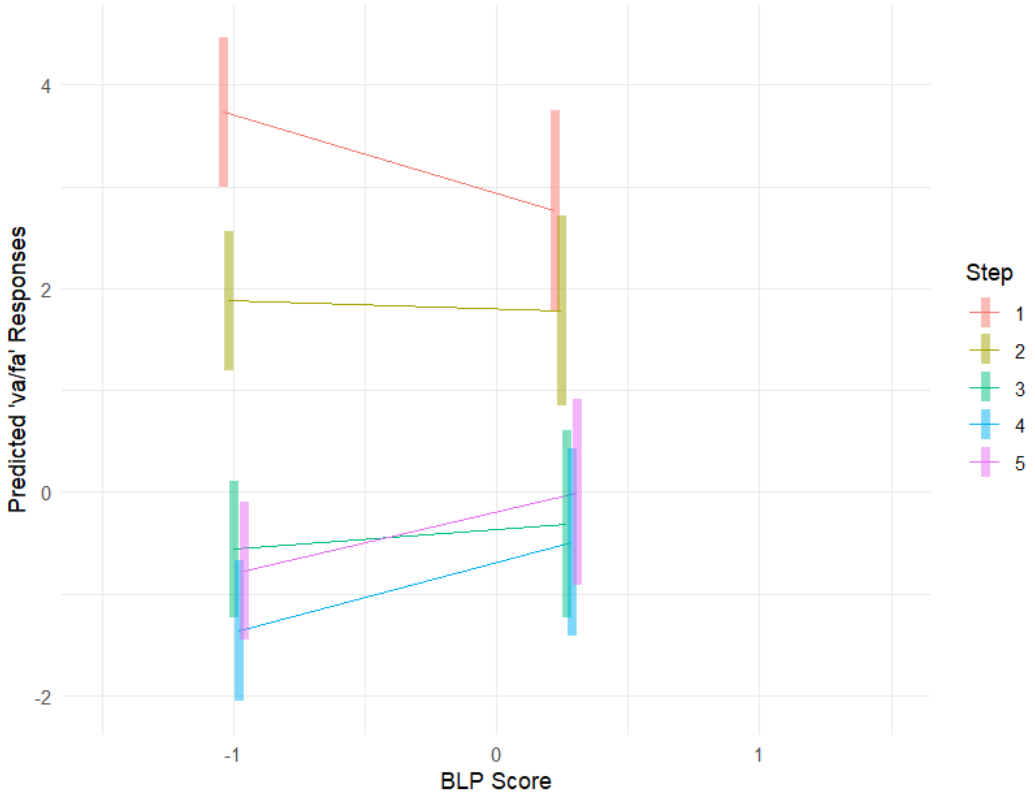

Figure 10: Estimated trends for the AV-va model depicting the interaction between Step and BLP Score

Results from the three-way contrast between BLP Score, AV Condition, and Step can be found in Table 16. Step 1 shows the greatest difference between AV condition with lower English dominance resulting in fewer 'va/fa' responses in the A-Only condition.

Among the AV conditions, Step 1 stood out, showing a larger estimate value compared to other AV contrasts at each Step. As seen in Table 16, the contrast between AV-va and A-Only at Step 1 yields a positive estimate value, indicating that as BLP score increases, the difference in 'va/fa' response between AV conditions also increases. In other words, individuals with weaker English dominance are more influenced by visual-va in perceiving 'va/fa' compared to A-Only perception. Perceiving 'va/fa' in the incongruent pairing of Step 5 and AV-va would be a case of experiencing the visual dominance illusion. For this incongruent pairing, there is no significant difference between AV-va and A-Only. However, when looking at just 'va' response (not 'va' and 'fa'), we do find a significant interaction between AV condition and BLP Score at Step 5 (B = 1.58, SE = 0.34, p = <0.001). This finding indicates that as English dominance decreased, the influence of visual-va in perceiving 'va' diminished across AV conditions. Individuals with greater English dominance experienced more of a boost in perceiving 'va' from visual-va compared to the A-Only condition. Lastly, the presence of visual-/va/ increases the likelihood of perceiving 'va/fa' at each Step compared to the A-Only condition, but the impact of visual information is the same across the BLP Score continuum, as there was no significant BLP effect between any pair of Steps in the AV-va condition.

Table 16: Estimated trends from the AV-va model of the three-way interaction between Step, AV Condition, and BLP Score

Step	AV Condition	Contrast	Estimate	SE	z.ratio	p-value
1	.	AV-va - A-Only	1.69	0.49	3.44	**0.01****
2	.	AV-va - A-Only	0.10	0.35	0.29	1.00
3	.	AV-va - A-Only	-0.32	0.28	-1.13	1.00
4	.	AV-va - A-Only	-1.06	0.34	-3.13	**0.04***
5	.	AV-va - A-Only	-0.26	0.29	-0.90	1.00
.	A-Only	Step 2 - Step 1	1.48	0.32	4.59	**<0.001****
.	A-Only	Step 3 - Step 1	1.97	0.35	5.61	**<0.001****
.	A-Only	Step 3 - Step 2	0.49	0.24	2.00	1.00
.	A-Only	Step 4 - Step 1	2.83	0.40	7.15	**<0.001****
.	A-Only	Step 4 - Step 2	1.35	0.31	4.41	**<0.001****
.	A-Only	Step 4 - Step 3	0.86	0.31	2.82	0.12
.	A-Only	Step 5 - Step 1	2.35	0.35	6.65	**<0.001****

.	A-Only	Step 5 - Step 2	0.87	0.25	3.50	**0.01****
.	A-Only	Step 5 - Step 3	0.39	0.25	1.52	1.00
.	A-Only	Step 5 - Step 4	-0.48	0.31	-1.54	1.00
.	AV-va	Step 2 - Step 1	-0.11	0.51	-0.21	1.00
.	AV-va	Step 3 - Step 1	-0.05	0.46	-0.10	1.00
.	AV-va	Step 3 - Step 2	0.06	0.38	0.16	1.00
.	AV-va	Step 4 - Step 1	0.08	0.46	0.16	1.00
.	AV-va	Step 4 - Step 2	0.18	0.38	0.49	1.00
.	AV-va	Step 4 - Step 3	0.12	0.30	0.41	1.00
.	AV-va	Step 5 - Step 1	0.40	0.46	0.86	1.00
.	AV-va	Step 5 - Step 2	0.51	0.38	1.33	1.00
.	AV-va	Step 5 - Step 3	0.44	0.30	1.48	1.00
.	AV-va	Step 5 - Step 4	0.32	0.30	1.07	1.00

4. Discussion

The objectives of this study were twofold: first, to investigate how Spanish and English dominance affects the perception of phonemes that differ across two languages, and second, to examine how visual speech information influences auditory perception in a language-dependent manner. To address these aims, we recruited English monolinguals and Spanish-English bilinguals to assess their perception /v/ and /b/, phonemes which are represented differently in Spanish and English, under both auditory and auditory-visual conditions.

In the A-Only condition, we found that lower English dominance resulted in greater perception of ‘ba,’ but only at Step 1. This was likely due to Step 1 being the original /va/ auditory stimulus, where at least English-dominant listeners would be unlikely to perceive it as ‘ba.’ As the continuum moved towards ‘ba’-like, the perception of 'ba' increased uniformly across different levels of language experience. For ‘va’ perception, the greatest difference was also at Step 1 (i.e., the original /va/ auditory stimulus), with weaker English dominance resulting in less ‘va’ perception, which is what we expected. However, as the continuum shifted more towards /ba/, we found that less English dominance results in an increase in ‘va’ perception. This shift in increased ‘va’ perception with less English dominance occurs at Steps 3, 4, and 5, when the auditory stimulus is most ambiguous. This goes against our initial expectations that weaker

English/stronger Spanish dominance would result in a bias towards hearing more ‘ba’ along the entire continuum. A possible explanation for this trend could be that under ambiguous phonetic situations, individuals with lower English dominance shift their phonetic boundaries to adjust for an English context. This shift in phonetic boundaries results in an overcompensation of perceiving ‘va,’ as /v/ is a phoneme that exists in English but not Spanish. Previous research has shown that in phonetic discrimination tasks based on voice onset time (VOT), Spanish-English bilinguals can exhibit an “English-like” phonemic boundary, even if Spanish was their first language (Garcia-Sierra et al., 2009). Bilinguals have also been shown to display an intermediate phonetic boundary, such that their average boundary of perception falls between that of the monolinguals of the two languages they know (Caramazza et al., 1973; Elman et al., 1977; García-Sierra et al., 2012). Spanish-English bilingual children (L1 Spanish, L2 English) have shown to shift towards English categorical boundaries after being exposed to Spanish-accented English speech (McDonald & Kaushanskaya, 2023). It is possible that due to the English context and the presentation of ambiguous auditory stimuli used in this study, lower English dominance led to a stronger shift towards English phonetics and an inhibition of Spanish phonemic knowledge.

In the AV condition, we found that for AV-ba, visual information influenced those with weaker English dominance greater than those with stronger English for perceiving ‘ba/ma/pa’ at Step 1. More balanced Spanish-English dominance resulted in perceiving illusory-ba at Step 1. This aligns with previous research suggesting that bilinguals are more susceptive to visual influence in speech perception (Chen & Hazan, 2007; Marian et al., 2018; Sekiyama, 1994). Interestingly, in the AV-ba condition, after Step 3, participants with weaker English dominance exhibited fewer 'ba/ma/pa' responses. One possible explanation for this phenomenon is the potential confusion between orthography and perception. While the phoneme /v/ does not phonetically exist in Spanish, it is present orthographically and is perceived as /b/. Consequently, individuals with weaker English dominance may have perceived 'ba' at these later steps but typed 'va,' as this would correspond to /ba/ in Spanish.

For AV-va, we saw that across the language dominance continuum, only Steps 1 and 4 showed significant differences between AV conditions for perception of ‘va/fa.’ Relative to A-only trials, listeners with weaker English dominance showed an increase in ‘va/fa’ perception when AV-va was present, agreeing with prior research that bilinguals have greater reliance upon visual cues. However, at Step 4, the opposite occurred, such that listeners with stronger English dominance perceived more of an increase in ‘va/fa’ between A-Only and AV-va conditions. We expected to see a more pronounced effect of visual context for bilingual individuals who had weaker English dominance. Although most prior research has indicated that bilinguals are more sensitive to visual speech cues, there is some research that shows the reverse. An audiovisual study looking at phonetic perception of a novel phonetic contrast found monolinguals to be more sensitive to visual speech cues, showing a stronger audiovisual benefit (Burfin et al., 2014). This suggests that the mechanism underlying visual speech differs between bilinguals and monolinguals.

Language context has been shown to play a significant role in bilinguals' processing of language (Casillas & Simonet, 2018; Garcia-Sierra et al., 2009; Marian & Spivey, 2003; Spivey & Marian, 1999). Given this, it is reasonable to posit that the English context in which this study was conducted substantially impacted the phonetic perception of bilingual participants, even under audiovisual conditions.

5. Conclusion

The current findings enhance our understanding of how bilinguals perceive phonemes that are treated differently across the languages they know, and how visual information impacts phonetic perception across language dominance. Moreover, these results shed light on the influence of visual information on phonetic perception across different levels of language dominance. However, a limitation of this study was the absence of a proper /ba/ audio recording. This shortfall makes it challenging to definitively determine whether the results accurately reflect how language experience modulates the perception of /v/ compared to /b/. Further steps include the use of different phonemes (such as /s/ and /z/, which are similar to /b/ and /v/), different language contexts, and looking at the neural mechanisms underlying perception of these phonemes in bilinguals.

Chapter 3: How top-down and bottom-up factors influence bilingual ability to restore degraded speech

1. Introduction

The human brain's ability to perceive and process speech in adverse and noisy environments is a complex and remarkable feat. One clear example of this capability is the phonemic restoration effect, an auditory phenomenon in which listeners are able to perceptually fill-in a degraded speech signal when there is an extraneous interruption (Samuel, 1981a; Warren, 1970). When a speech segment is removed and replaced with noise, oftentimes listeners fill-in the missing speech segment and perceive the speech signal as continuous through the noise.

The phonemic restoration effect has been found to be modulated and enhanced by both bottom-up and top-down effects (Bashford et al., 1996; Groppe et al., 2010; Samuel, 1981a, 1997; Shahin & Miller, 2009; Sivonen et al., 2006). Top-down contextual factors such as the presence of visual information, lexical identity, and sentential context have all been reported to aid in this illusory perception (Groppe et al., 2010; Shahin & Miller, 2009; Sivonen et al., 2006). For example, the presence of congruent visual cues (i.e., the talker's mouth movements) facilitates enhanced perception of degraded speech as continuous, particularly as the duration of missing speech segments increases (Shahin & Miller, 2009). Furthermore, the lexical properties of the degraded item itself (e.g., word length and whether it is a word or nonword) also play a role in the extent of the illusion, such that longer words produce greater illusion and nonwords yield less illusion (Samuel, 1985a).

Additionally, bottom-up factors, such as the type of extraneous sound used and the acoustic properties of the phoneme replaced, have also been found to impact phonemic restoration (Bashford et al., 1996; Samuel, 1981a, 1981b). For example, if a fricative is removed and replaced with white noise, individuals are more likely to successfully perceive the speech as continuous than if a vowel was removed (Samuel, 1981a). However, the opposite occurs when a pure tone is used instead of white noise (Samuel, 1981a). Overall, the phonemic restoration effect suggests that there are complex interactions between top-down and bottom-up processes in auditory speech perception.

Understanding the neural underpinnings of the phonemic restoration effect provides deeper insight into the mechanisms underlying this perceptual phenomenon. Research has investigated how the illusion of continuous perception manifests in the brain. When speech is perceived as continuous, the N1-P2 auditory-evoked potential (AEP), an event related potential (ERP) that responds to sound onsets and offsets, is shown to have a reduction in amplitude, compared to when the illusion fails (Shahin et al., 2012). This reduction in AEP amplitude reflects a suppression of auditory cortex to interruption boundaries (e.g., onsets/offsets), which in turn is thought to facilitate filling-in of the missing phoneme and thus illusory continuity perception (Shahin et al., 2012). Previous research has also found a reduction in theta band power and theta phase-locking in response to illusory filling-in (Riecke et al., 2009; Shahin et al., 2012), following interruption boundaries. Taken together, these studies demonstrate that successful

phonemic restoration is mediated by inhibition of the auditory cortical response to interruption boundaries, reflected in suppressed AEPs and theta oscillatory activity.

It is important to note that the majority of research on the phonemic restoration effect involves populations of native speakers of English. There is a notable gap in the literature concerning both the subjective experience of the illusion and its underlying neural mechanisms in bilinguals. It is known that bilinguals can understand their second language (L2) speech just as well as native speakers of that language in quiet conditions (Florentine, 1985; Mayo et al., 1997; Takata & Nábělek, 1990), but bilinguals perform less accurately at speech recognition when listening to their L2 (or less dominant language) in the presence of background noise (Bsharat-Maalouf & Karawani, 2022; Fricke, 2022; M. L. G. Lecumberri et al., 2010; Mayo et al., 1997). The question arises about bilinguals' ability to perceive degraded speech in the case of the phonemic restoration effect. This effect reflects the strength of phonemic encoding as a function of both bottom-up and top-down modulations. One study investigated how Japanese-English bilinguals restore degraded speech compared to English monolinguals, focusing on the context of words and pseudowords as well as the acoustic properties of the removed phoneme (Ishida & Arai, 2016). Using nasals (present in both English and Japanese) and liquids (present only in English), no difference in phoneme restorability was observed between the language groups. Additionally, lexical context significantly influenced restorability only for English monolinguals, who restored words more effectively than pseudowords. This effect was not observed in Japanese-English bilinguals. This study showed how the integration of bottom-up and top-down cues depends upon the listeners' language proficiency (Ishida & Arai, 2016).

It has been suggested that bilingualism results in enhanced sound perception through the utilization of top-down cognitive mechanisms (Krizman et al., 2012; Levi, 2018). However, most of the research investigating bilingualism and top-down effects on speech perception has concentrated on comparing task performance of bilingual individuals with those of monolinguals rather than exploring *how* specific top-down cues affect bilingual performance in particular tasks. Some research has examined how bilingual individuals utilize top-down cues, such as language context and linguistic knowledge/experience, to facilitate speech perception. For instance, top-down cues of language context significantly shift bilinguals' phonetic boundaries during phonetic categorization tasks, favoring the context of the language they are using (Casillas & Simonet, 2018; Garcia-Sierra et al., 2009; Wig & García-Sierra, 2021). However, to our knowledge, there has been no research examining how the top-down effect of attention modulates bilingual speech perception in comparison to monolinguals.

Therefore, the goals of the current study were: 1) to examine how language experience interacts with lexical context, the language specificity of the missing phoneme, and the role of attention during phonemic restoration and 2) to understand how language experience influences phonemic restoration and its underlying neurophysiological mechanisms. Specifically, this study investigated heritage speakers of Spanish (HSS) in the US (i.e., individuals who were raised speaking Spanish at home in a society dominated by English) due to unique phonological differences between Spanish and English. Spanish phonology, unlike English, does not have a distinction between the

phonemes /b/ and /v/. In Spanish, the phoneme /v/ maps onto /b/ which is why often, many native Spanish speakers replace /v/ sounds in English words with /b/ (e.g., *volleyball* is pronounced like *bolleyball*). Because of this distinct phonological difference between the two languages, the question arises of how HSS respond to degraded speech in which /b/ and /v/ phonemes are removed. To address these aims, we recorded EEG data while presenting HSS and English monolinguals with degraded words and pseudowords in active and passive tasks. This approach aimed to assess how top-down cues from lexical knowledge influence the restoration of degraded speech. The inclusion of both active and passive tasks further facilitated exploration of top-down information. Moreover, the degraded speech stimuli were generated by substituting either the phoneme /b/ or /v/ with white noise, aiming to investigate how phoneme type affects participants' susceptibility to the illusion.

Thus, the current study aimed to achieve the following using EEG: 1) uncover how lexical context impacts bilinguals' responses to degraded speech, 2) examine how bilinguals perform in phonemic restoration under conditions with and without top-down attentional input, and 3) investigate how the bottom-up properties of the removed phoneme influence phonemic restoration in bilinguals. Additionally, from a behavioral perspective, we sought to investigate how bilingualism (specifically one's language dominance) interacts with lexical context and the language-specificity of the missing phoneme.

Referring to the theoretical framework I proposed in Chapter 1, I assume that in their L2, bilinguals tend to pay more attention to irrelevant auditory stimuli, such as noise and interruptions in speech, because of weaker phonemic encoding, than monolinguals (see Fricke, 2022; Lecumberri & Cooke, 2006; Lecumberri et al., 2010). Additionally, bilinguals' sensitivity to low-level acoustic cues should be broader, resulting in stronger encoding of these cues. This broader filter exacerbates their already weakened phonemic encoding. Because of this weakened phonemic encoding and sensitivity to bottom-up cues, we expect HSS to be less likely to fill in missing /v/ phonemes than English monolinguals, resulting in larger N1-P2 AEPs and increases in theta phase-locking for the bilinguals. We don't expect to see differences in N1-P2 AEPs and oscillatory activity between groups when /b/ is the missing phoneme, since both English and Spanish contain /b/. Additionally, my framework states that bilinguals rely more on lexical and semantic information, making selective attention crucial for accurately filling in missing parts of speech. Compared to the active task, in a passive task, HSS should show larger N1-P2 amplitudes and greater theta phase-locking, due to a lack of top-down attentional information. The same is expected for word and pseudoword stimuli, respectively.

2. Materials and Methods

2.1. Participants

Thirty-six individuals participated in this study. However, four participants were excluded due to technical issues during data collection or unclean data, resulting in usable data from thirty-two participants (>18 years of age, M = 20.5 years, SD = 4.7 years; 19 females, 13 males). Of these thirty-two participants, sixteen were classified as English monolinguals and sixteen as Spanish-English bilinguals (see Table 1). These language

group categorizations were formed based on participant responses to the Bilingual Language Profile (BLP), a language questionnaire that generates a Language Dominance Score from a participant's responses. (Birdsong et al., 2012). The BLP provides a quantitative measure of language dominance, where values near -1 indicate English dominance, values of 0 indicate balanced English-Spanish dominance, and values near 1 indicate Spanish dominance. All participants self-reported normal hearing, normal or corrected vision, and no language deficits.

Table 1: Demographic information of participants.

	Monolingual	**Bilingual**
N	16	16
Age	20.13 (4.06)	20.88 (5.32)
English Age of Acquisition	0 (0)	2.44 (2.37)
Spanish Age of Acquisition		0.44 (1.75)
BLP Score	-0.78 (0.18)	-0.19 (0.18)

Participants were recruited via an internal recruiting system of the University of California, Merced or through fliers, and provided written consent prior to participation. All experimental protocols were approved by the Institutional Review Board (IRB) of the University of California, Merced, and all methods were carried out in accordance with the guidelines and regulations of the IRB of the University of California, Merced and in accordance with the Declaration of Helsinki.

2.2. Stimuli

The study consisted of two main categories of stimuli: words and pseudowords spoken by a female talker (mean f0 = 205 Hz). The stimuli were spoken by a female native English speaker and recorded using a RØDE NTG2 microphone at a sampling rate of 48k Hz, in an acoustically shielded room. All auditory stimuli were normalized in Adobe Audition (Adobe, Inc., San Jose, CA) to the same sound intensity and were presented at ~62.4 dBA sound pressure level. Each word/pseudoword category consisted of words that contained either the phoneme /b/ or /v/ in the second or third syllable. There were 10 trisyllabic words and 3 5-syllabic words per category of word/pseudoword and phoneme. Pseudowords were created by replacing at least two phonemes, other than the target phoneme, from each lexical condition in the Word category, to ensure that the pseudowords followed English phonotactics. Each word/pseudoword had three recordings, or tokens, for a total of 156 stimuli (10 trisyllabic words + 3 pentasyllabic words x 2 word categories x 2 phonemes x 3 tokens = 156).

To elicit the phonemic restoration illusion, each word/pseudoword had the target phoneme of either /b/ or /v/ removed from either the second or third syllable and replaced with silence. To do so, we first located the sample in each original audio file at which the target phoneme was initially produced using Adobe Audition. Starting at each located sample for each stimulus, 125 ms of the audio file was silenced by using custom code in MATLAB (MathWorks, Natick, MA, USA). This silent gap was then replaced with 125 ms of Gaussian white noise at a constant dBW-power across all stimuli.

2.3. Procedure

Participants sat in a sound attenuated booth (IAC Acoustics, Naperville, IL) about 127 cm from a 22-inch computer monitor with two external Yamaha HS8 speakers on either side, located at ±45˚ relative to the listener. Each participant underwent three blocks during the task: Active (EEG), Passive (EEG), and Behavioral block. EEG data were recorded during both the Active and Passive blocks, but not during the Behavioral block. The order of the Active and Passive blocks was counterbalanced across participants, while the Behavioral block was consistently the ultimate task. Stimuli were presented using Presentation software v.20.3 (Neurobehavioral Systems, Inc., Berkeley, CA, USA). Prior to the start of all the blocks, participants were told they would be presented with words, which might sound like real English words or might seem similar to English words.

In the Active block, participants were presented with a randomized stream of all the word and pseudoword stimuli, which contained tri- and 5-syllabic items (156 stimuli). Intertrial intervals (ITI) ranged between 3000 and 4100 ms, in steps of 100 ms. The participants' task was to pay attention to the stream and respond by pressing the space bar with their left index finger when they heard a 5-syllabic word or pseudoword. By instructing participants to identify the target items, the task required the use of top-down information (i.e., attention and lexicosemantic knowledge). Participants were also instructed to restrain from counting syllables out loud or on their fingers to prevent motor artifacts in the data. A light gray background with a black fixation cross was presented on the monitor while stimuli were presented.

In the Passive block, the same stimuli were used but randomized differently, and participants were told to pay attention to a compilation of silent movies as the stream played in the background. These were three short, animated videos in which the characters did not speak and there were no subtitles. As a result, assuming their attention was focused on the movie, participants should not be relying as heavily on top-down knowledge and should be depending more on bottom-up information during perception of the speech stimuli.

Similar to the Active and Passive blocks, the Behavioral block consisted of the same stimuli presented in a randomized order (excluding the 5-syllabic stimuli). Each stimulus was presented, and participants' task was to listen carefully and determine whether they heard the stimulus as "interrupted" or "continuous." As quickly as possible, subjects pressed their left index finger on the left or right arrow key when they perceived the stimulus as interrupted or continuous, respectively. Prior to the start of the Behavioral

block, participants were played two exaggerated examples of what a continuous or interrupted perception would sound like. The examples used the word *impulsive*, a word not used during the actual block. Similar to the creation of the experimental stimuli, either 125 ms or 300 ms of white noise was inserted into the word "impulsive," with the noise onset occurring at the start of the /s/ phoneme. This was done to ensure participants understood the difference between a continuous and interrupted percept.

2.3. EEG Acquisition

EEG data were recorded using a 64-channel EEG system (BioSemi Active Two system, 10–20 Ag-AgCl electrode). Data were sampled at a rate of 1024 Hz, and an antialiasing low-pass filter was applied. No reference electrode was selected since BioSemi systems record reference-free. Electrode offsets, relative to the Common Mode Sense electrode, for all channels were set to < 20 μV.

2.4. EEG Preprocessing

EEG analyses were conducted using ERPLAB (Lopez-Calderon & Luck, 2014), EEGLAB (Delorme & Makeig, 2004), and FieldTrip (Oostenveld et al., 2011) toolboxes and in-house MATLAB code. Preprocessing steps included: 1) Downsampling each subjects' data to 512 Hz. 2) Each dataset was epoched (segmented) from -1.75 to 1.2 s around the onset of each word/pseudoword and baselined to the entire epoch (mean removed). 3) Independent component analysis (ICA) was conducted on each individual dataset, producing 64 ICA components. 4) ICA components that reflected ocular artifacts were removed, and bad channels were interpolated using EEGLAB's spherical interpolation. 5) Individual datasets were average referenced. Datasets were further processed in two ways: to produce files suitable for ERPs and files suitable for time-frequency activity.

2.4.1. ERPs

Individual files were filtered between 0.1 and 30 Hz (zero-phase Butterworth, fourth order) and re-epoched around the onset of the noise interruption from -1.5 to 2.1 s. Datasets were re-baselined to the pre-stimulus onset period (-1 to -0.9 s) to account for the duration of the word or pseudoword before the noise interruption. This adjustment ensures that the baseline corresponds to the silent portion of the inter-trial interval (ITI) before the word or pseudoword is played. Then, trials with amplitude shifts greater than 150 μV at any channel were removed. For the sake of simplicity, the data were sorted according to the following conditions: Active, Passive, Word, Pseudoword, /b/ removed, and /v/ removed. The mean number of remaining trials and ranges for each condition across subjects was as follows: Active and Passive (avg = 115, [76 127], Word and Pseudoword (avg = 115, [62 125]), /b/ and /v/ (avg = 114, [65 122]). Finally, trials within each stimulus condition were averaged, creating separate ERP files for each condition.

2.4.2. Time Frequency

Individual files were re-epoched around the onsets of the noise interruption from -1.5 to 1.5 s. Epochs with amplitude shifts larger than 200 μV in any channel were rejected, and trials were sorted into the conditions mentioned above. The remaining mean number of trials and ranges across subjects were: Active (avg = 110, [74 123]), Passive (avg = 108, [78 120]), Word (avg = 110, [80 121]), Pseudoword (avg = 109, [88 121]), /b/

(avg = 109, [84 121]), and /v/ (avg = 110, [90 121]). Event-related spectral perturbation (ERSP, spectral power) and inter-trial phase coherence (ITPC, phase-locking) spectrograms were generated for each condition, channel, and subject using the timef.m function of the EEGLAB toolbox. We examined activity between -1500 and 1500 ms relative to the onset of the noise interruption for the 4 – 50 Hz frequency range. The analysis used a sinusoidal wavelet-based Discrete Fourier transform (DFT) of the time-domain signal (512-ms Hanning window, 10 ms and 1 Hz steps) with two cycles at the lowest frequency, increasing linearly to 5 cycles at the highest frequency. As frequency increases, the window size linearly decreases. For the ERSP analysis, post-stimulus activity was baselined to the pre-interruption onset in the period spanning from -1500 to -875 ms.

2.5. Statistical Analyses

2.5.1. Behavioral

Logfiles of participants' responses were transferred to Excel spreadsheets, which were then parsed using custom MATLAB code. Participants' responses were categorized according to whether they responded "continuous" or "interrupted." A table containing information about the lexical condition (word or pseudoword), the phoneme removed (/b/ or /v/), the response, and response time (RT) for each trial was created for each participant. Response times were converted into milliseconds and subsequently logarithmically transformed to normalize the data distribution.

For statistical analysis, we created a binomial distribution of the response data, such that a "continuous" response was coded as 1 and "interrupted" responses were coded as 0. Using R Studio (R version 4.3.2) and the lme4 package (Bates et al., 2015), we fitted Generalized Linear Mixed-Effects Models (GLMMs) to the data using logistic regression, since we coded the response data in a binomial manner.

First, we conducted a mixed effects binomial linear regression on the data to examine whether language experience (monolingual or bilingual), lexical condition (word or pseudoword) and phoneme (/b/ or /v/ removed) predict subjects' perception of the illusion. Using the function glmer, the formula was "continuous" perception ~ Language Experience * Lexical Condition * Phoneme Removed + 1|SubjectID. We also utilized the bobyqa optimizer.

A similar mixed effects linear regression was run on the RT data to examine whether language experience, lexical condition, phoneme, and response type (perceived "continuous" or "interrupted") predict subjects' response time. Using the function lmer, the formula was RT ~ Language Experience * Lexical Condition * Phoneme Removed * Response + 1|SubjectID. For both the RT and response models, the emmeans package (Lenth, 2024) was used to further investigate any interaction effects, and multiple comparisons of these post-hoc tests were controlled for via Bonferroni correction.

2.5.2. EEG

ERP amplitude differences between bilinguals and monolinguals for each condition were examined using the nonparametric cluster-based permutation test (CBPT;

Maris & Oostenveld, 2007; Oostenveld et al., 2011) as implemented in the FieldTrip toolbox. The CBPT considers activity at all time-points and channels. For the ERP data, the CBPT identifies significantly different EEG amplitudes between the groups using independent-samples t-tests at each data point (alpha = 0.05). These significant points are grouped into continuous clusters based on their spatial and temporal adjacency. Within each cluster, the t-statistics for each time-channel sample are summed. To create the null distribution, this process is repeated for each resampling of the data, for a total of 2000 permutations (Monte Carlo simulation). For each permutation, the largest cluster t-statistic sum is recorded. The summed t-statistics for each cluster in the real data are then compared against this null distribution, using an alpha value of 0.05. Clusters that exceed this alpha value are considered to be significantly different between the groups being compared (i.e., Monolinguals vs. Bilinguals for Active, Passive, Word, Pseudoword, /b/, and /v/).

To statistically test for group differences in the time-frequency data (i.e., the spectral power and phase-locking data) for each condition, a nonparametric permutation test was used at each time and frequency bin of the channel spectrograms. All channels were included in the resampling process to control for multiple comparisons (refer to Chau et al., 2004, for a detailed description of this method). The data were normalized into decibels (dB) by computing the ratio of post-stimulus to pre-stimulus values, then taking the logarithm. At each time-frequency point and for each channel, an independent-samples t-test was performed. Null distributions were generated from a 325-ms period before the acoustic stimulus (from -1500 to -875 ms, with a 10 ms temporal resolution and 1 Hz spectral resolution), using maximum values obtained from repeated resampling of the data (2000 permutations, significance level set at p = 0.05). The t-statistics for each time-frequency point at each channel in the real data were then compared against this null distribution, using an alpha value of 0.05. Data points that exceed this alpha value are considered to be significantly different between the groups being compared.

3. Results

3.1. Behavioral

The primary binomial linear regression analysis was performed to examine how language experience, lexical condition, and phoneme removed impact perception of the illusion. For this model, Word, Monolingual, and /b/ were set as the referent levels. Results from the model predicting subjects' perception of the illusion (i.e., "continuous" response) can be seen in Figure 1 and Table 2.

Table 2: Results from binomial linear regression model predicting perception of the illusion, with factors of lexical condition, language group, and phoneme removed.

Fixed Effect	Estimate	SDE	Statistic	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	-0.07	0.28	-0.27	0.79	-0.63	0.47

Pseudoword	1.24	0.14	9.11	**<0.001*** **	0.97	1.5
Bilingual	0.54	0.4	1.36	0.18	-0.24	1.32
/v/	0.07	0.12	0.53	0.6	-0.18	0.31
Pseudoword: Bilingual	-1.23	0.18	-6.72	**<0.001*** **	-1.59	-0.87
Pseudoword:/v/	-0.26	0.19	-1.4	0.16	-0.63	0.12
Bilingual:/v/	0.11	0.17	0.65	0.52	-0.23	0.45
Pseudoword: Bilingual:/v/	0.13	0.26	0.52	0.61	-0.37	0.63

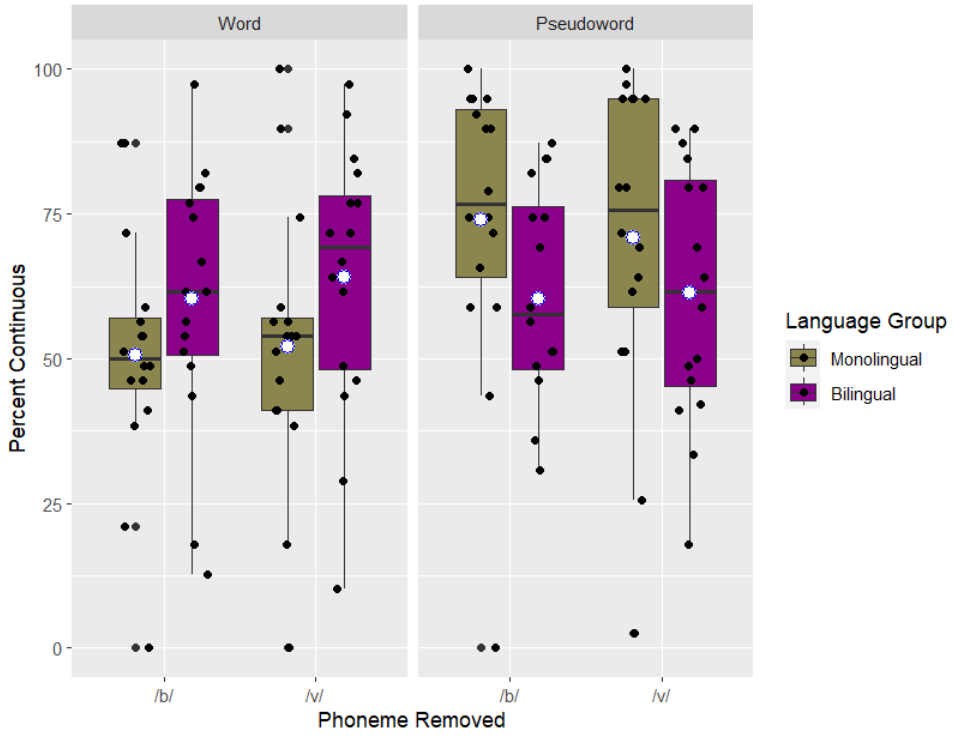

Figure 1:Three-way interaction between lexical condition, language group, and phoneme removed. Y-axis indicates the average frequency of perceiving degraded speech as continuous.

The model revealed a significant main effect of Pseudoword with a positive estimate value, suggesting that compared to words, pseudowords predict greater perception of continuity (see Figure 2). Additionally, there was also a negative significant two-way interaction between Language Group and Lexical Type. To further understand the two-way interaction, we conducted post-hoc pairwise contrasts and plotted these results using the emmeans library in R. As shown in Table 3 and Figure 2 below, only the contrast between Pseudoword and Word for English monolinguals yielded a statistically significant result, indicating that English monolinguals were significantly more likely to perceive Pseudowords as continuous compared to Words. However, the bilinguals did not show a significant difference in continuity perception between Words and Pseudowords, and no group differences were observed within the Word and Pseudoword conditions.

Table 3: Post-hoc pairwise contrast between language group and lexical condition.

Language Group	Lexical Condition	Contrast	Estimate	SE	z ratio	p-value
English Monolingual		Pseudoword-Word	1.1	1.0	11.61	**<0.001 *** **

Spanish-English Bilingual		Pseudoword-Word	-0.06	0.09	-0.67	1.0
	Word	Bilingual - Monolingual	0.59	0.39	1.53	0.5
	Pseudoword	Bilingual - Monolingual	-0.57	0.39	-1.46	0.58

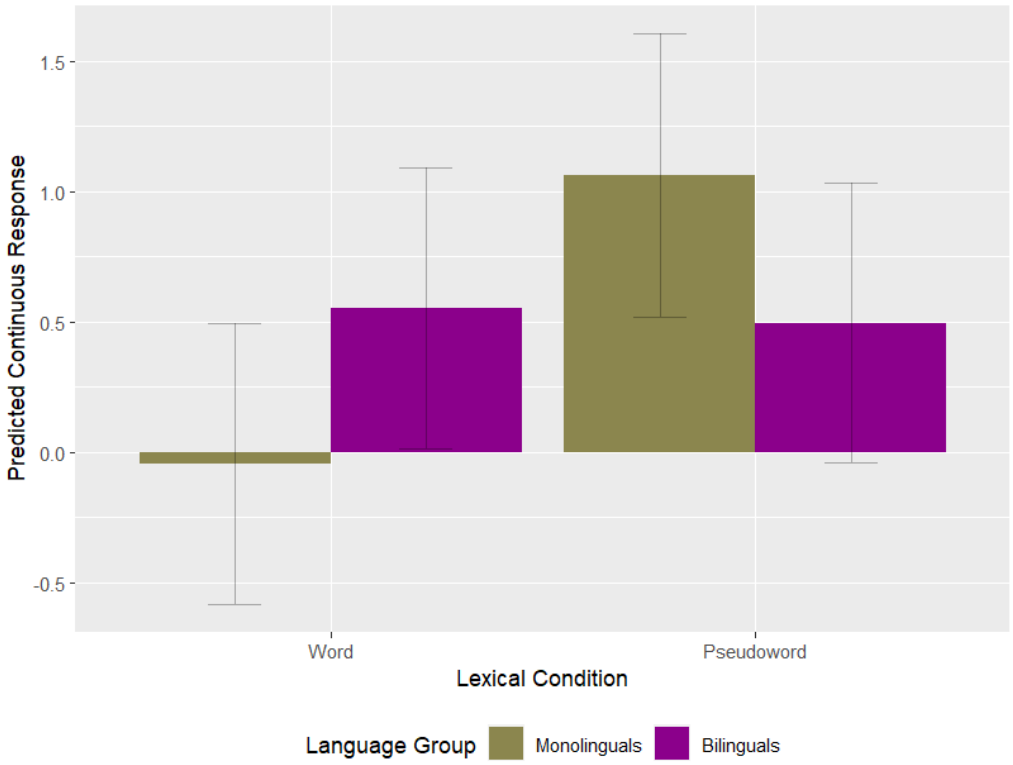

Figure 2: Estimated marginal means from the continuous perception model depicting the interaction between lexical condition and language group.

The second linear regression analysis was performed to examine if lexical condition, language group, the phoneme removed, and perception of stimuli (as continuous or interrupted) influenced response time. For this model, Word, Monolingual, and /b/ were set as the referent levels for the outcome measure. Results from the model predicting subjects' response time can be seen in Table 4.

Table 4: Linear regression model predicting RT, with factors language group, lexical condition, phoneme removed, and response.

Fixed Effect	Estimate	SDE	Statistic	p-value	95% CI (LL)	95% CI (UL)
(Intercept)	8.16	0.03	264.98	0	8.09	8.22
Interrupted	-0.02	0.01	-1.52	0.13	-0.05	0.01
Pseudoword	0.01	0.01	1.1	0.27	-0.01	0.04
Bilingual	-0.09	0.04	-2.07	**0.05***	-0.18	0
/v/	0.01	0.01	0.95	0.34	-0.01	0.04
Interrupted:Pseudoword	0.05	0.02	2.11	**0.04***	0	0.09
Interrupted:Bilingual	0.06	0.02	2.77	**0.006****	0.02	0.1
Pseudoword:Bilingual	0.05	0.02	2.77	**0.006****	0.02	0.09
Interrupted:/v/	-0.03	0.02	-1.29	0.2	-0.07	0.01
Pseudoword:/v/	-0.01	0.02	-0.42	0.68	-0.04	0.03
Bilingual:/v/	0.01	0.02	0.56	0.58	-0.03	0.05
Interrupted:Pseudoword:Bilingual	-0.1	0.03	-3.23	**0.001****	-0.16	-0.04
Interrupted: Pseudoword:/v/	0.06	0.03	2.03	**0.04***	0	0.12
Interrupted:Bilingual:/v/	-0.02	0.03	-0.66	0.51	-0.08	0.04
Pseudoword:Bilingual:/v/	-0.01	0.03	-0.5	0.61	-0.06	0.04
Interrupted:Pseudoword:Bilingual:/v/	0.01	0.04	0.16	0.87	-0.08	0.09

Firstly, the model revealed a significant main effect for the Bilingual group, indicated by a negative estimate value. This suggests that bilinguals, compared to monolinguals, have shorter response times overall. However, two significant two-way interactions (one between the perception of speech and group, and another between lexical type and group) as well as a significant three-way interaction involving group, complicate the interpretation of this main effect. These interactions suggest that the relationship between bilingualism and response times is influenced by how speech is perceived and the type of stimuli. Post-hoc pairwise contrasts delve further into these interactions. Starting with the interaction between language group and response, pairwise comparisons yielded no significant contrasts (see Figure 3). Despite the presence of a two-way interaction in the original model, these pairwise comparisons did not reveal significant contrasts, indicating that the interaction effect might be more nuanced or that additional factors influenced response time.

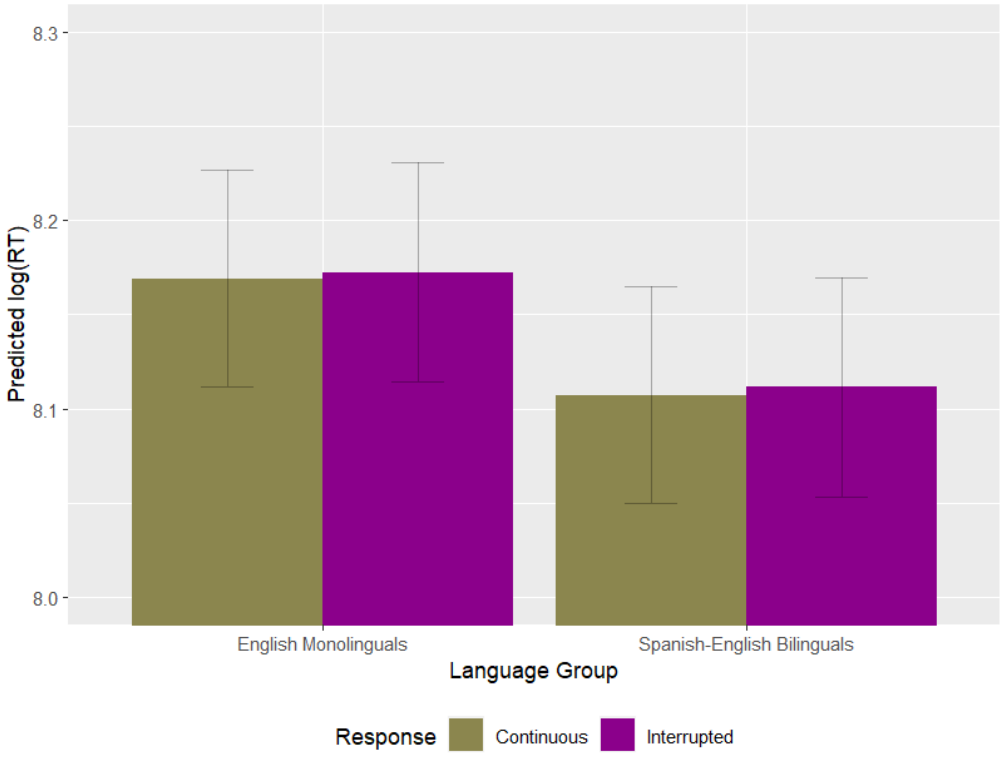

Figure 3: Estimated marginal means from the RT model depicting the interaction between response and language group.

The interaction between language group and lexical condition are visualized in Figure 4. Post-hoc pairwise comparisons yielded no significant differences in response times between bilinguals and monolinguals for either words or pseudowords. Both monolingual and bilingual groups exhibited significantly longer response times for pseudowords compared to words. Monolinguals showed a slightly larger effect (monolingual: $B = 0.0491$, $SE = 0.00765$, $p < .0001$; bilingual: $B = 0.0465$, $SE = 0.00735$, $p < .0001$). This indicates that pseudowords consistently elicit longer response

times across both language groups, with a slightly more pronounced effect for monolinguals.

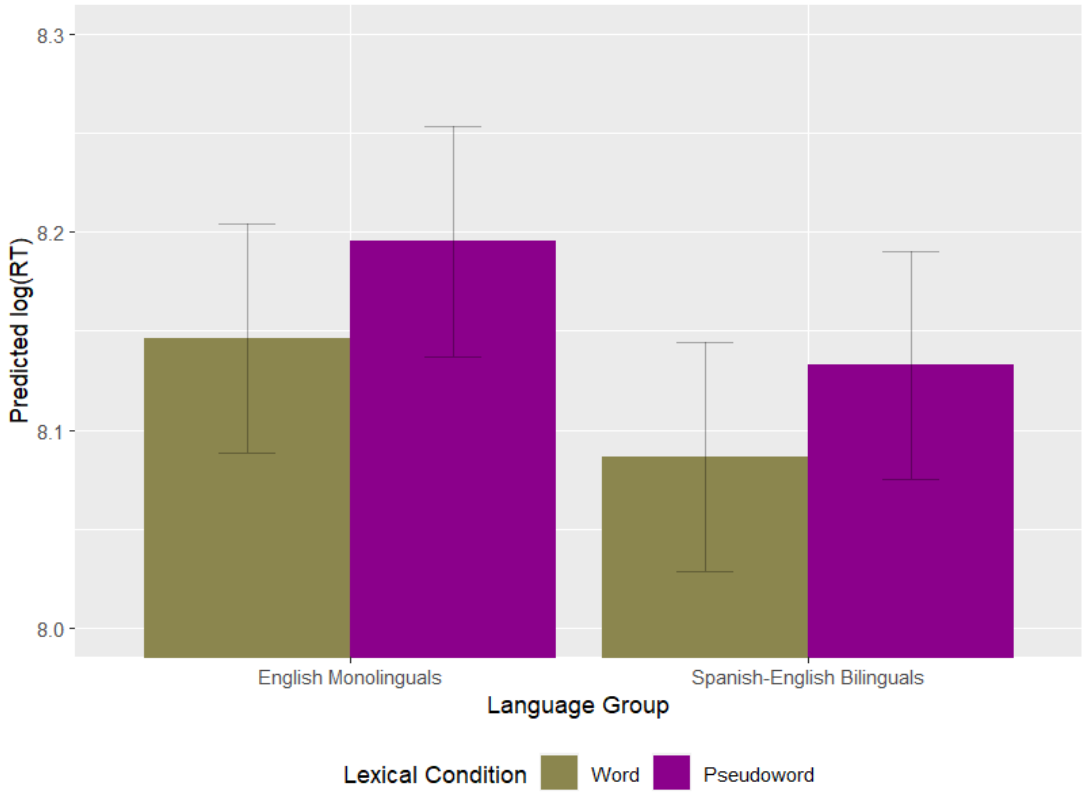

Figure 4: Estimated marginal means from the RT model depicting the interaction between lexical condition and language group

The RT model also revealed a significant two-way interaction between response and lexical condition (see Figure 5). Pairwise contrasts revealed the following significant comparisons: for continuous responses, there was a statistically significant difference ($B = 0.0329$, $SE = 0.00656$, $p < 0.0001$) when comparing pseudowords to words. For interrupted responses, there was also a statistically significant difference ($B = 0.0628$, $SE = 0.00856$, $p < 0.0001$) when comparing pseudowords to words. However, when participants perceived the stimulus as interrupted, there was a larger difference in RTs between pseudoword and word stimuli than when they perceived the stimulus as continuous.

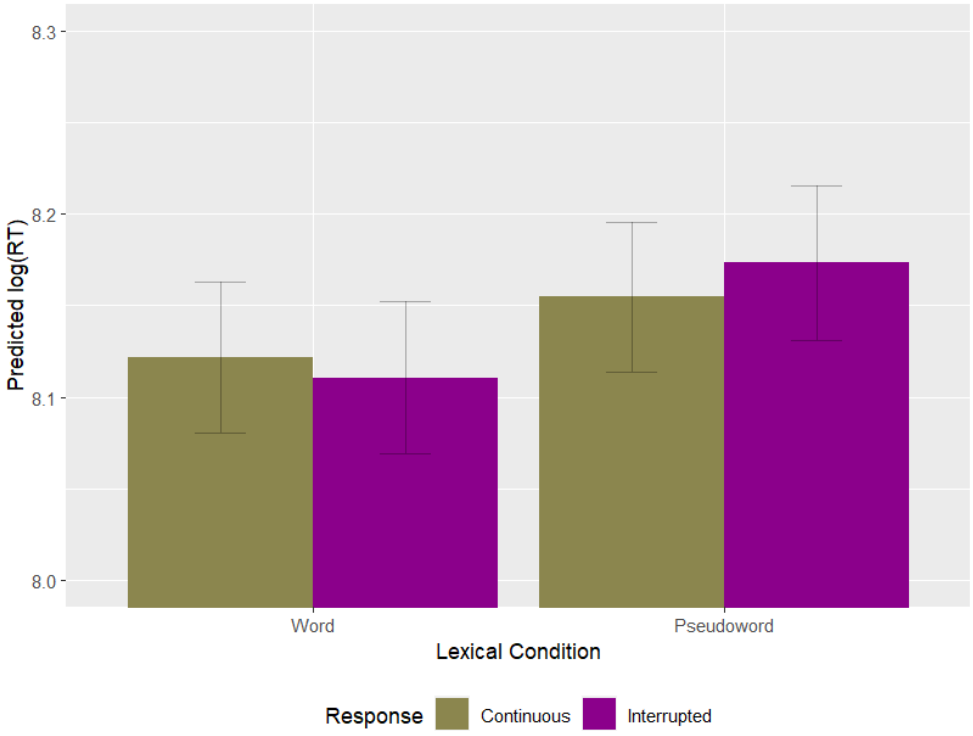

Figure 5: Estimated marginal means from the RT model depicting the interaction between response and lexical condition.

Lastly, there were two three-way interactions between 1) response, lexical condition, and language group (Figure 6) and 2) response, lexical condition, and phoneme removed (Figure 7). Starting with the first of the two, post-hoc pairwise comparisons revealed the following: For the monolinguals, interrupted pseudowords had longer RTs than continuous pseudowords (B = 0.04, SE = 0.01, p = 0.008), but this pattern is reversed in the word condition (B = -0.04, SE = 0.01, p = 0.01). However, we observed a different pattern of results for the bilinguals, such that interrupted pseudowords were not significant in RT from continuous pseudowords (B = -0.05, SE = 0.01, p = 1.00). Similarly, bilinguals also did not show significant differences in RT between interrupted and continuous words (B = 0.01, SE = 0.01, p = 1.00). Bilinguals also exhibited longer RTs for continuous pseudowords compared to continuous words (B = 0.06, SE = 0.01, $p < 0.001$), and for interrupted pseudowords and interrupted words (B = 0.04, SE = 0.01, p = 0.02). However, monolinguals only exhibited significantly longer RT for interrupted pseudowords compared to interrupted words (B = 0.09, SE = 0.01, $p < 0.001$).

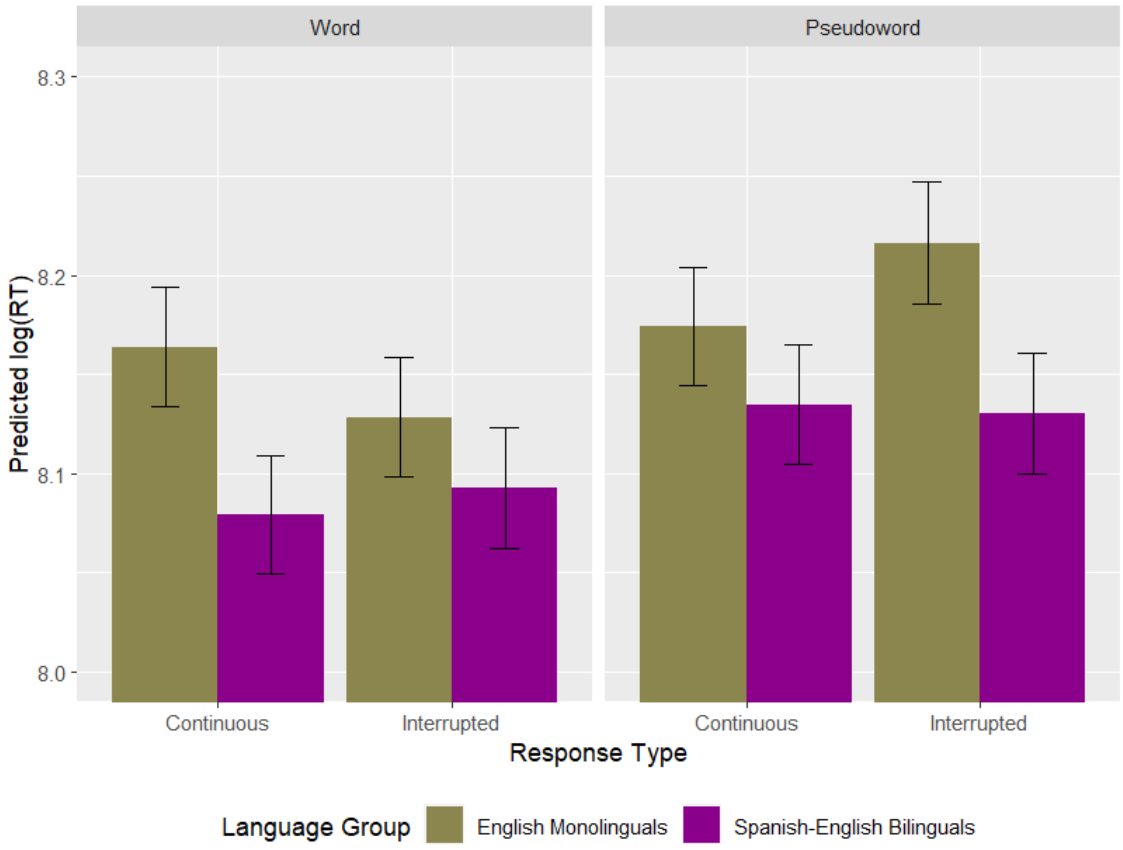

Figure 6: Estimated marginal means from the RT model depicting the three-way interaction between response, language group, and lexical condition.

Lastly, the three-way interaction between response, lexical condition, and phoneme removed can be seen in Figure 7. Pseudowords where a /v/ had been removed resulted in significantly longer RTs when perceived as interrupted compared to continuous (B = 0.03, SE = 0.01, p = 0.03), but this is not seen for pseudowords where a /b/ had been removed. Pseudowords with a /b/ removed that were perceived as continuous had significantly longer RTs than a continuous word with a /b/ removed (B = 0.04, SE = 0.01, p = <0.001); however, this was significant difference was not found for continuous pseudowords compared to continuous words with a /v/ removed. When /b/ or /v/ were removed from interrupted pseudowords, response times were longer compared to interrupted words, with the removal of /v/ resulting in slightly longer response times (/b/: (B = 0.04, SE = 0.01, p = 0.02); v: (B = 0.09, SE = 0.01, p < 0.001).

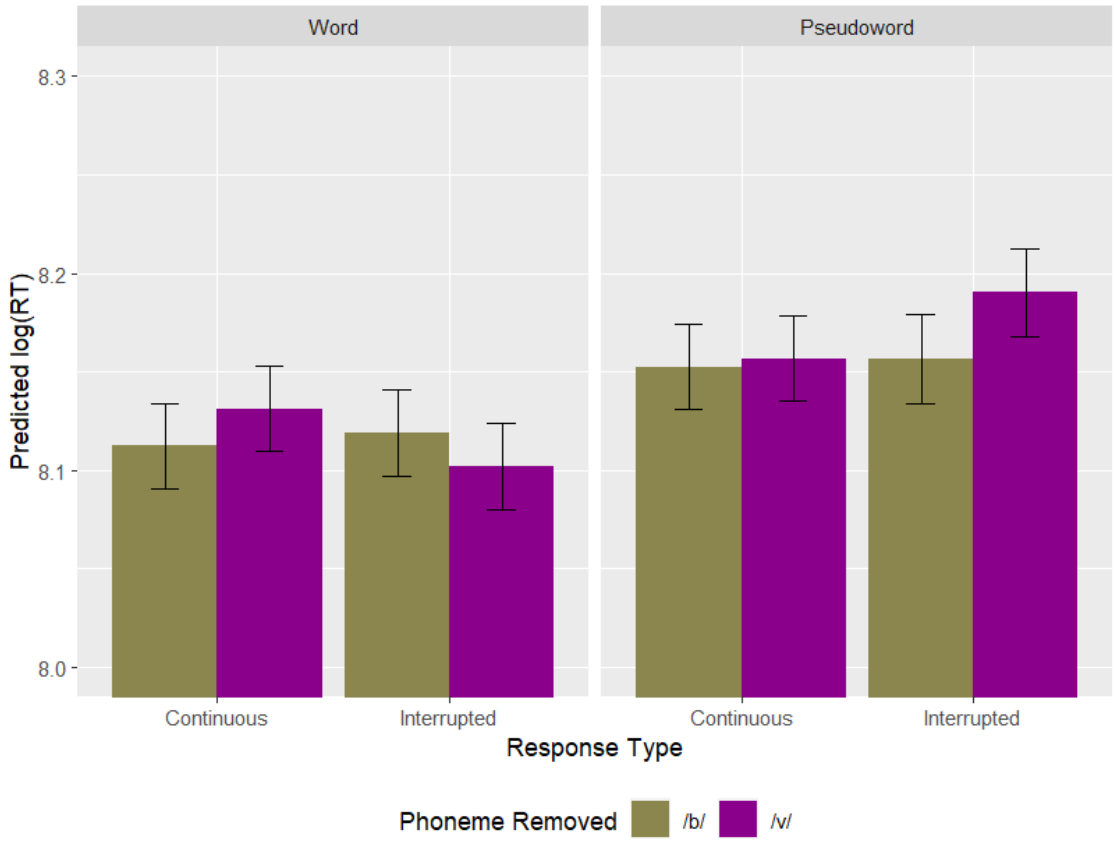

Figure 7: Estimated marginal means from the RT model depicting the three-way interaction between response, phoneme removed, and lexical condition.

3.2. ERPs

The cluster-based permutation tests conducted on ERP data revealed no significant group effects in ERP amplitude for any condition.

3.3. Spectral Power

We conducted permutation tests to determine the time and frequency range that showed significant group (Bilingual vs. Monolingual) differences in spectral power separately for each condition (Active, Passive, Word, Pseudoword, /b/, and /v/).

For the Active condition, Figure 8 depicts the significant differences in spectral power between bilinguals and monolinguals, as well as the group difference topographies of alpha (9-15 Hz, at -95 ms and 455 ms) and beta (12-30 Hz at 605 ms) power. We observed greater alpha suppression for bilinguals than monolinguals approximately 100 ms before the onset of the noise interruption; this alpha effect is observed over bilateral parietal scalp sites. Additionally, there was a subsequent increase in alpha suppression for bilinguals around 450 ms, also observed over bilateral parietal scalp sites. The results also revealed a significant reduction in beta power around 600 ms after the onset of the interruption in bilateral central electrodes.

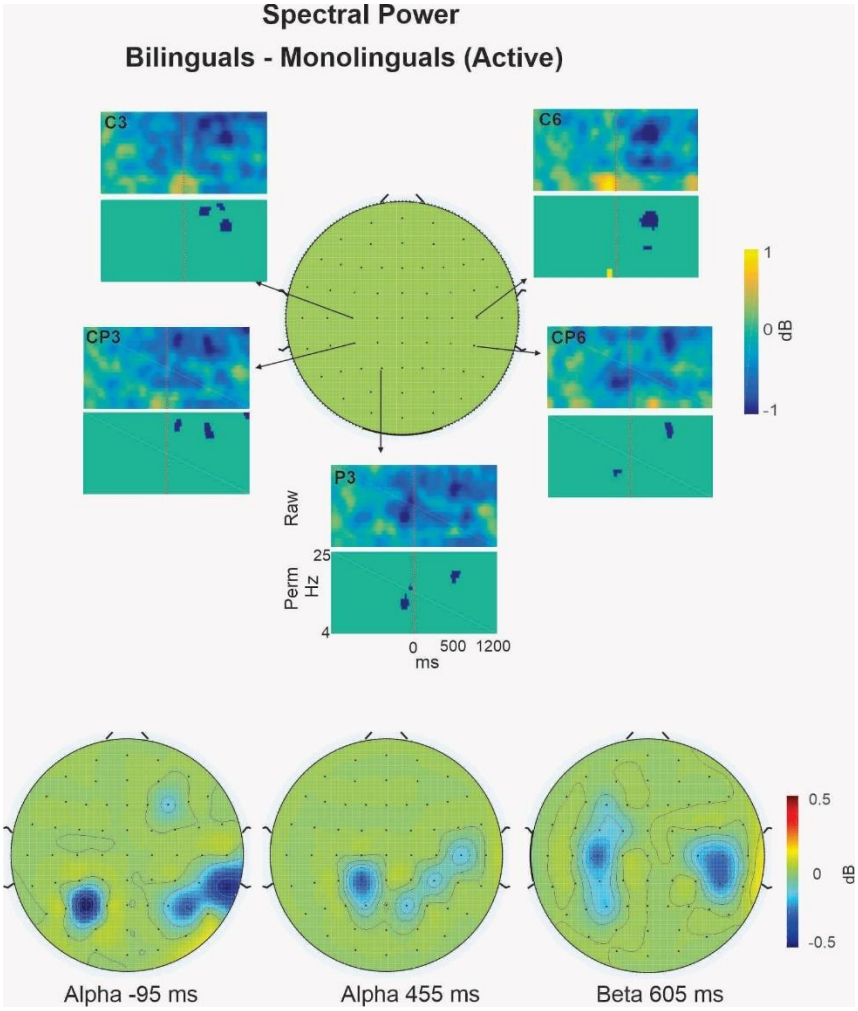

Figure 8: Spectral Power for Active Condition. The top spectrogram shows the raw power difference between groups; the bottom spectrograms indicate the data points where there was a significant difference between groups ($p < 0.05$). Time 0 ms is the onset of the noise interruption. The topographies show the raw group difference in power at peak time points.

For the Passive condition, we observed significantly greater theta power (4-8 Hz, ~0 ms) for bilinguals compared to monolinguals, specifically around the onset of the noise interruption (see Figure 9). The topography illustrates this theta power group difference, which is especially strong in fronto-central electrodes.

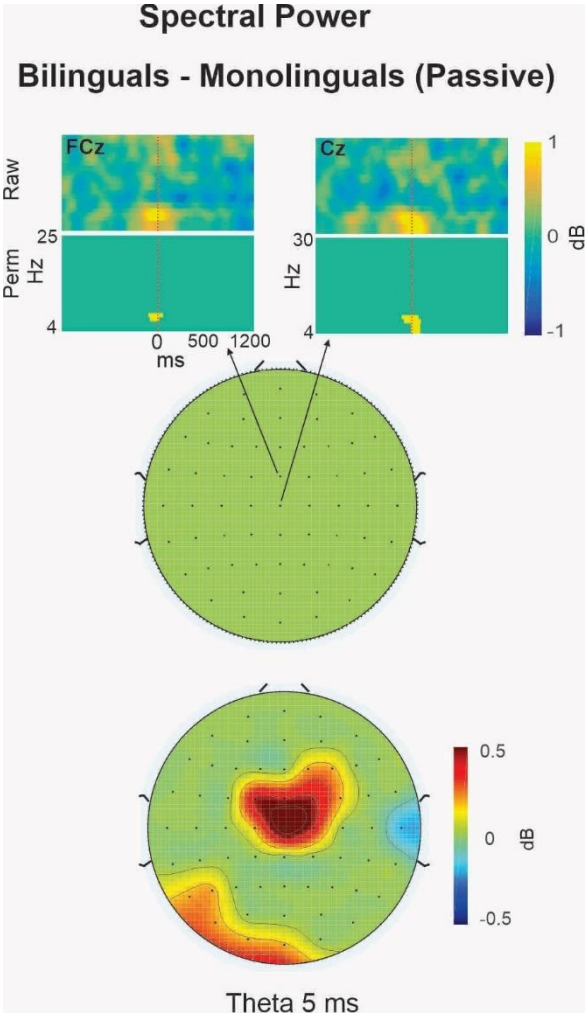

Figure 9: Spectral Power for Passive Condition

In the Word condition (see Figure 10), greater alpha power suppression for bilinguals than monolinguals was observed at several time points (i.e., coinciding with the onset of the noise interruption, as well as at approx. 450 and 950 ms following the noise interruption). The initial alpha effect at the noise interruption onset is then followed by bursts of gamma power enhancement (> 30 Hz, at 385 ms, 400 ms, and 710 ms) that temporally alternate with bursts of greater alpha suppression for bilinguals. The topographies show that pre-onset alpha power group difference is predominantly right-lateralized in parieto-occipital and temporal electrodes. Gamma enhancement is observed to be left-lateralized over temporal scalp sites, whereas later alpha suppression is evident in left centroparietal electrodes.

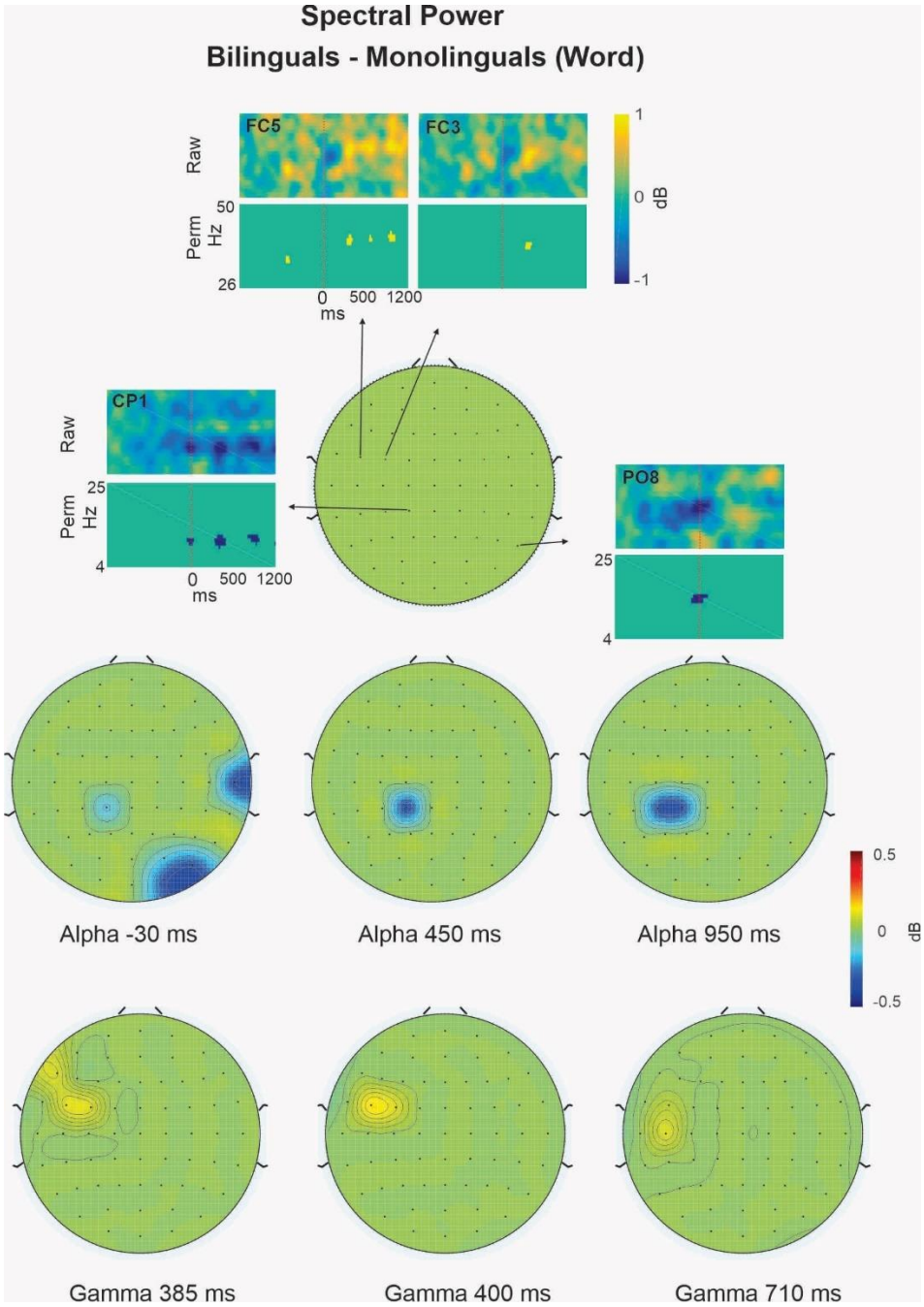

Figure 10: Spectral Power for Word Condition

Figure 11 shows the significant group differences in spectral power for the Pseudoword condition. There is significantly greater theta enhancement for bilinguals than monolinguals around the onset of the interruption. The topographies display this theta power group difference in fronto-central electrodes. Additionally, relative to monolinguals, the bilinguals had enhanced gamma power in bilateral frontal channels, occurring between approximately 400 and 750 ms after noise interruption onset.

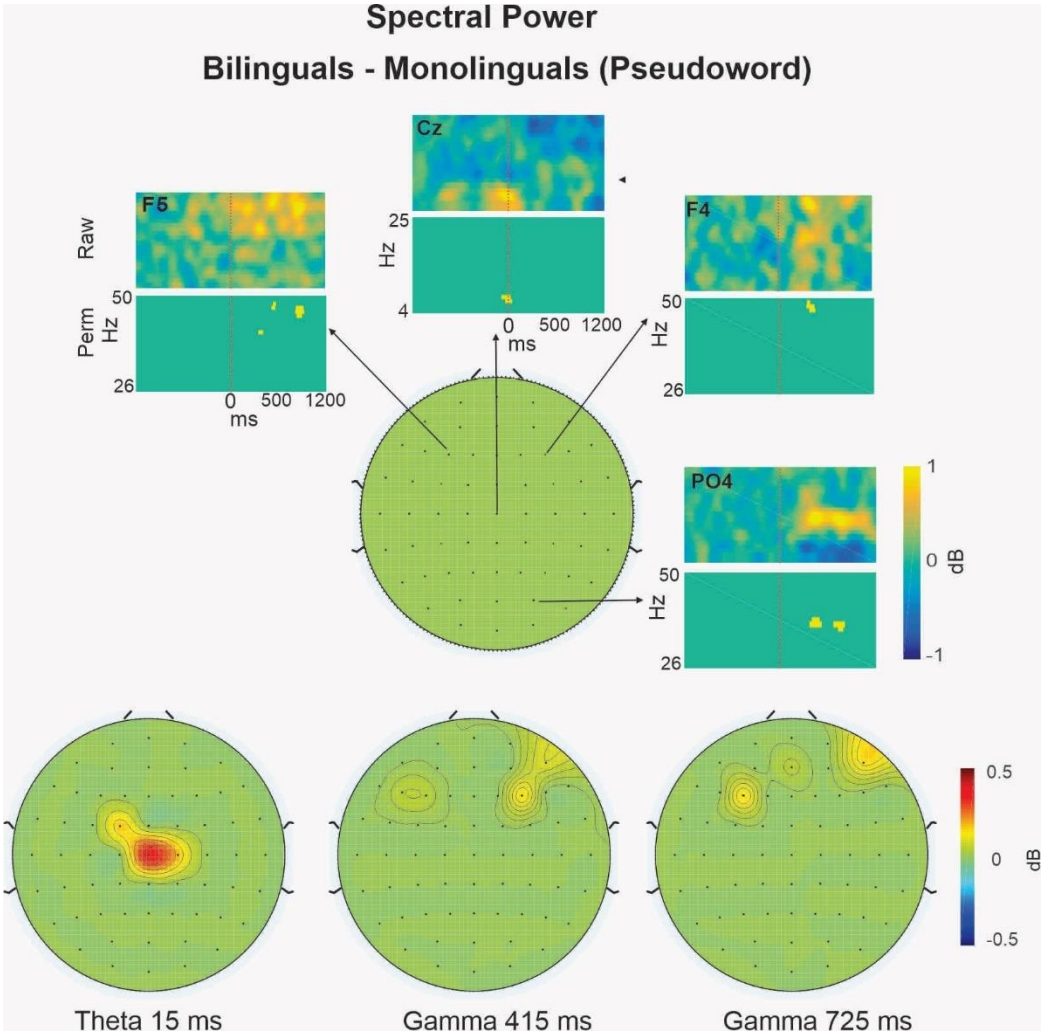

Figure 11: Spectral Power for Pseudoword Condition

Similarly, for the /b/ condition, significant group differences in theta power can be observed approximately around the interruption onset, as highlighted in the spectrograms (see Figure 12). Specifically, the bilinguals exhibited greater theta power at the noise interruption onset than the monolinguals. The topography shows that this theta power group difference predominantly occurs over central sites.

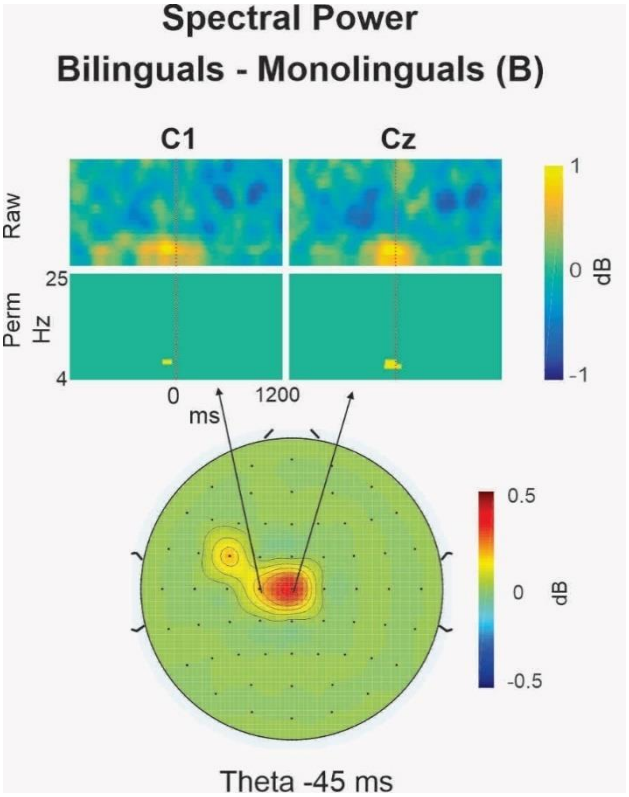

Figure 12: Spectral Power for /b/ Condition

Lastly, for the /v/ condition, we observed significant group differences in theta, alpha, and gamma power (see Figure 13). First, the spectrograms show spectral power differences of alpha suppression (in right temporo-parietal electrodes) and increased theta power (in fronto-central channels) around the onset of the interruption; these effects were stronger for bilinguals than monolinguals. This is then followed by enhanced gamma power around 300 ms (in left-lateralized frontal channels) for bilinguals and ending with enhanced alpha power around 900 ms for bilinguals. This alpha enhancement for bilinguals is evident in left central electrodes (700 ms) and in right parieto-occipital electrodes (900 ms).

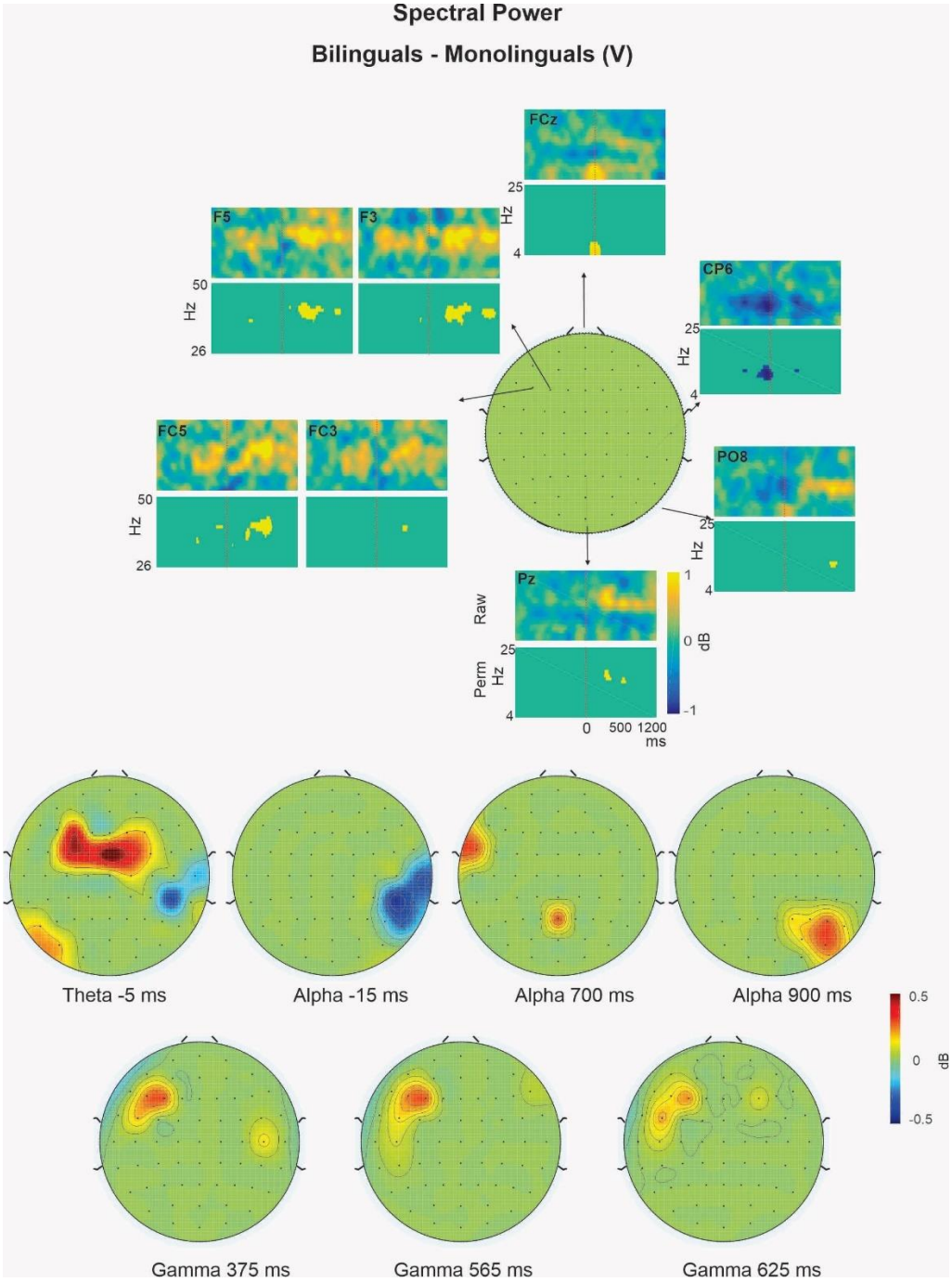

Figure 13: Spectral Power for /v/ Condition

3.4. Phase-Locking

We conducted permutation tests to determine the time and frequency range that showed significant group (Bilingual vs. Monolingual) differences in phase-locking separately for each condition (Active, Passive, Word, Pseudoword, /b/, and /v/).

Starting with the Active condition (Figure 14), relative to the monolinguals, there is an increase in theta phase-locking in the bilingual group at the onset of the interruption. The topography conveys enhanced theta phase-locking in occipital, left-parietal, and right-frontal electrodes for bilinguals.

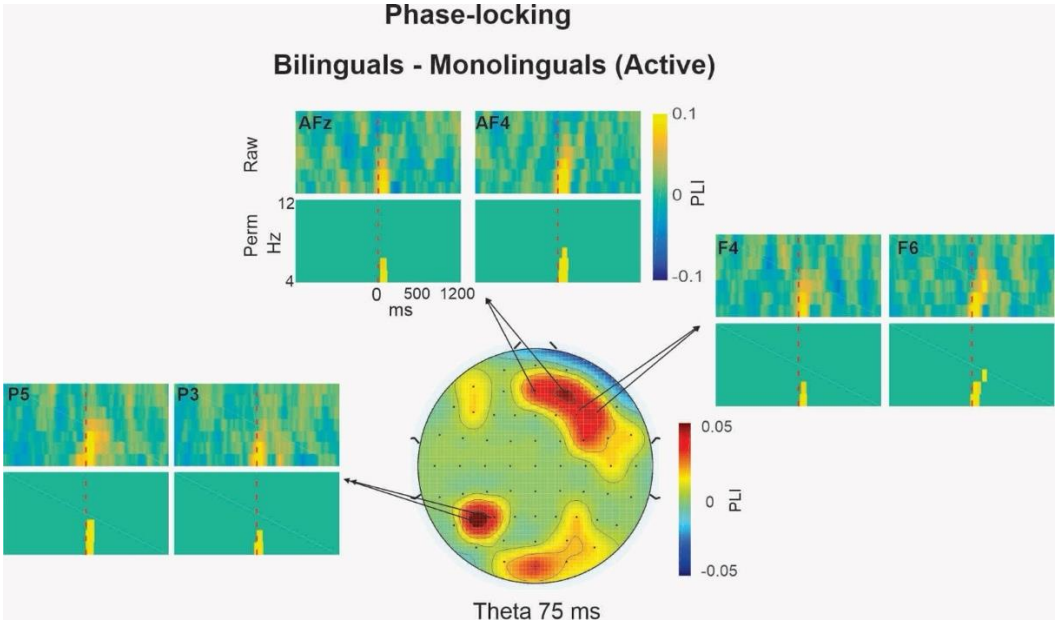

Figure 14: Phase-locking for Active Condition The top spectrogram shows the raw phase-locking difference between groups; the bottom spectrograms indicate the data points where there was a significant difference between groups ($p < 0.05$). Time 0 ms is the onset of the noise interruption. The topographies show the raw group difference in phase-locking at peak time points.

The Passive condition (see Figure 15) displays an increase in theta/alpha phase-locking for bilinguals around 200 ms after the noise interruption. The topography illustrates enhanced theta in bilinguals in left-lateralized centrotemporal channels.

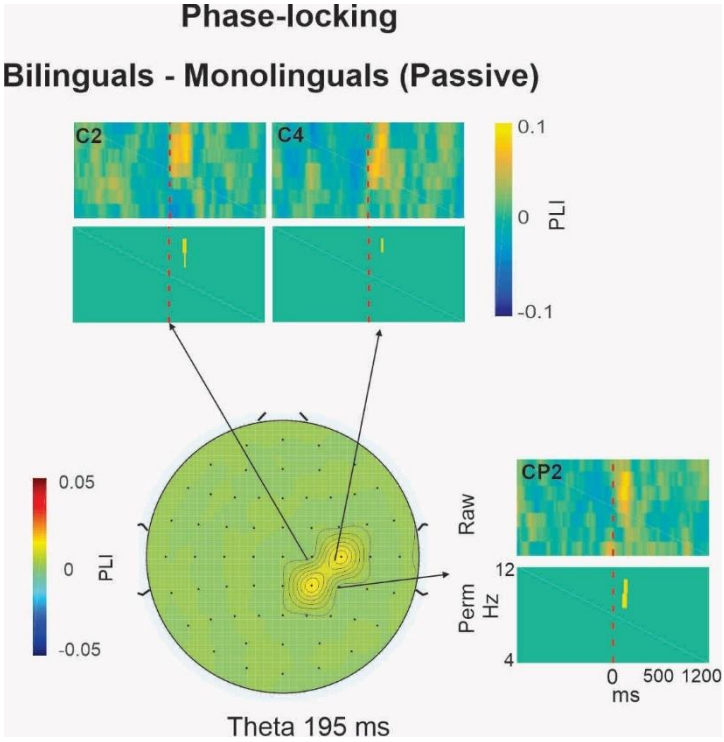

Figure 15: Phase-locking for Passive Condition

The /b/ condition (see Figure 16) once again displays an increase in theta phase-locking for bilinguals at the onset of the interruption and shortly thereafter. Bilinguals exhibited enhanced theta in left frontotemporal and right parietal channels at the onset, followed by greater theta in right frontotemporal and parietal channels, as seen in the topographies.

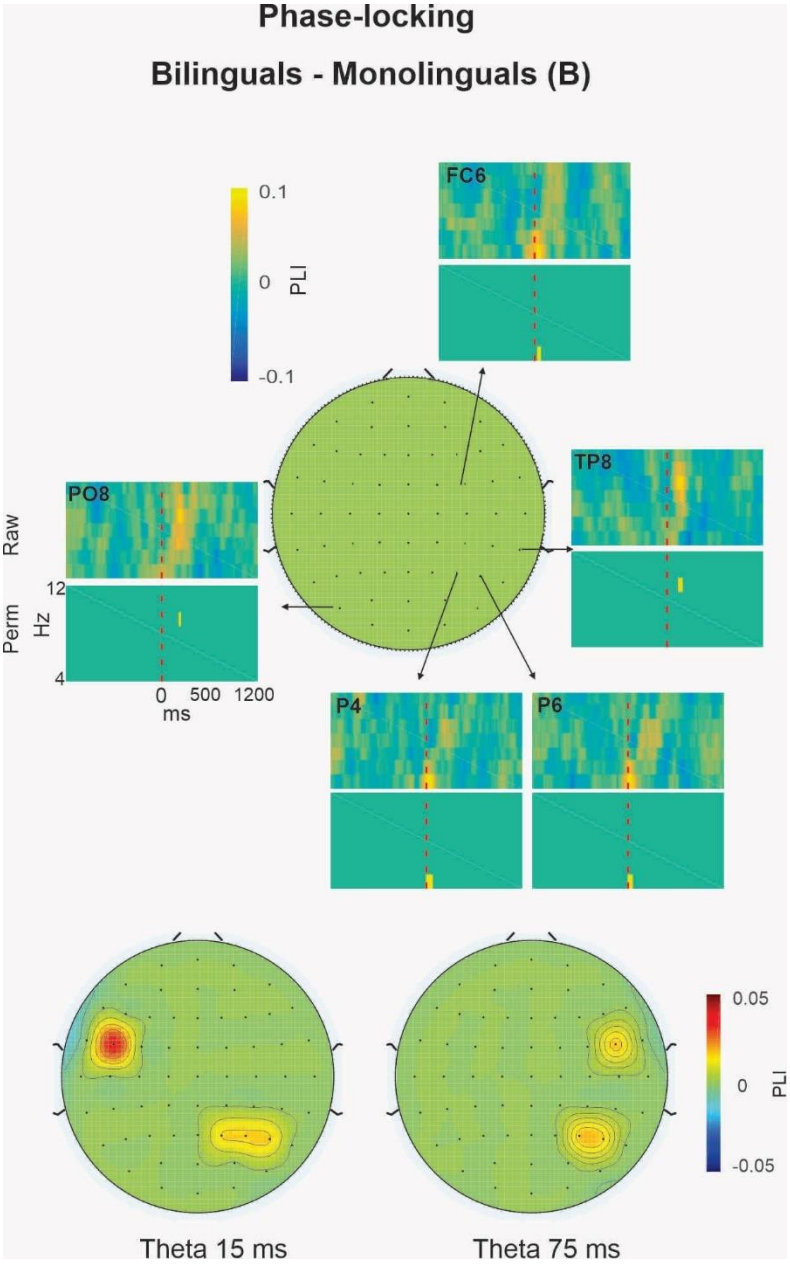

Figure 16: Phase-locking for /b/ Condition

Similarly, in the /v/ condition (see Figure 17), there is an increase in theta phase-locking for bilinguals at the onset of the interruption. Just as with the /b/ topography, the /v/ topography shows enhanced theta phase-locking in right frontotemporal and right parietal channels at the onset of the interruption for bilinguals (at approx. 75 ms). Notably in the Active condition, the theta phase-locking effect had a similar topography, whereby bilinguals also exhibit enhanced theta phase-locking in left-parietal and right-frontal electrodes.

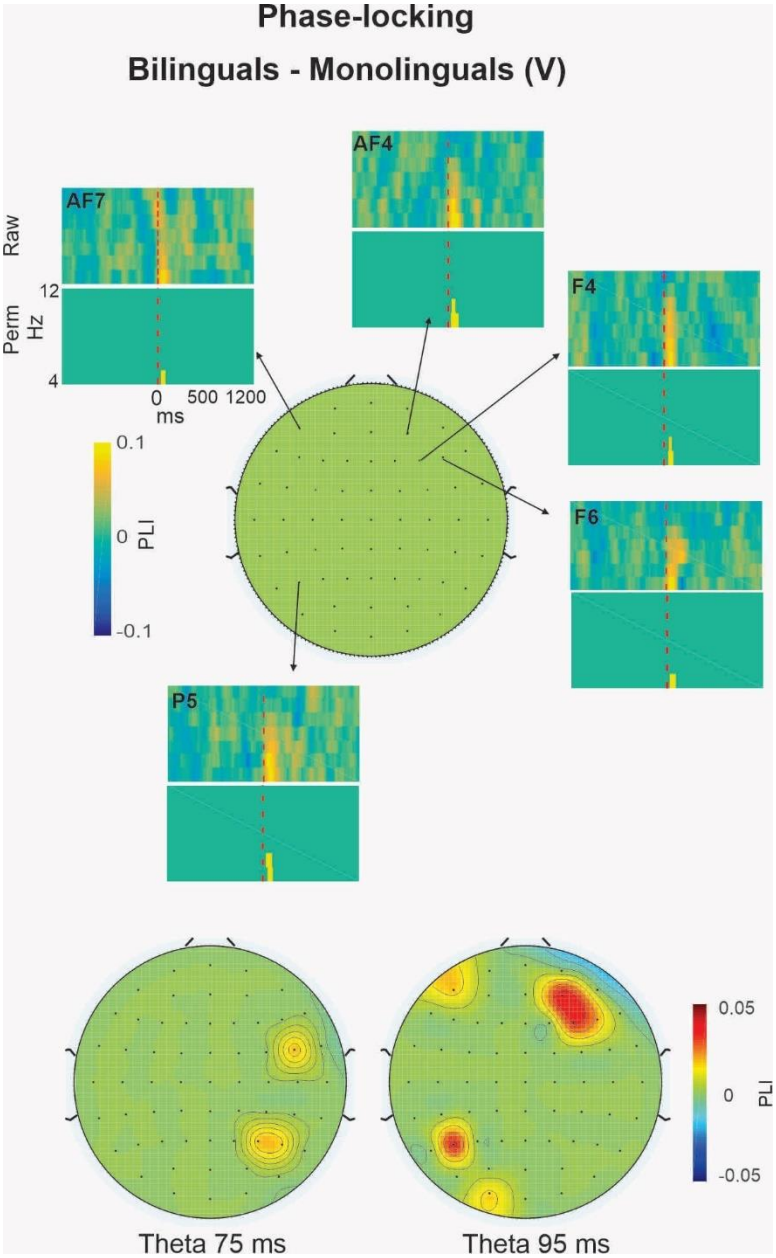

Figure 17: Phase-locking for the /v/ Condition

In the Word condition, there was an increase in theta-phase locking for bilinguals at the onset of the interruption. The topographies for this condition were more complex, with bilinguals exhibiting greater theta activity in left frontal, right frontotemporal, and parietooccipital channels.

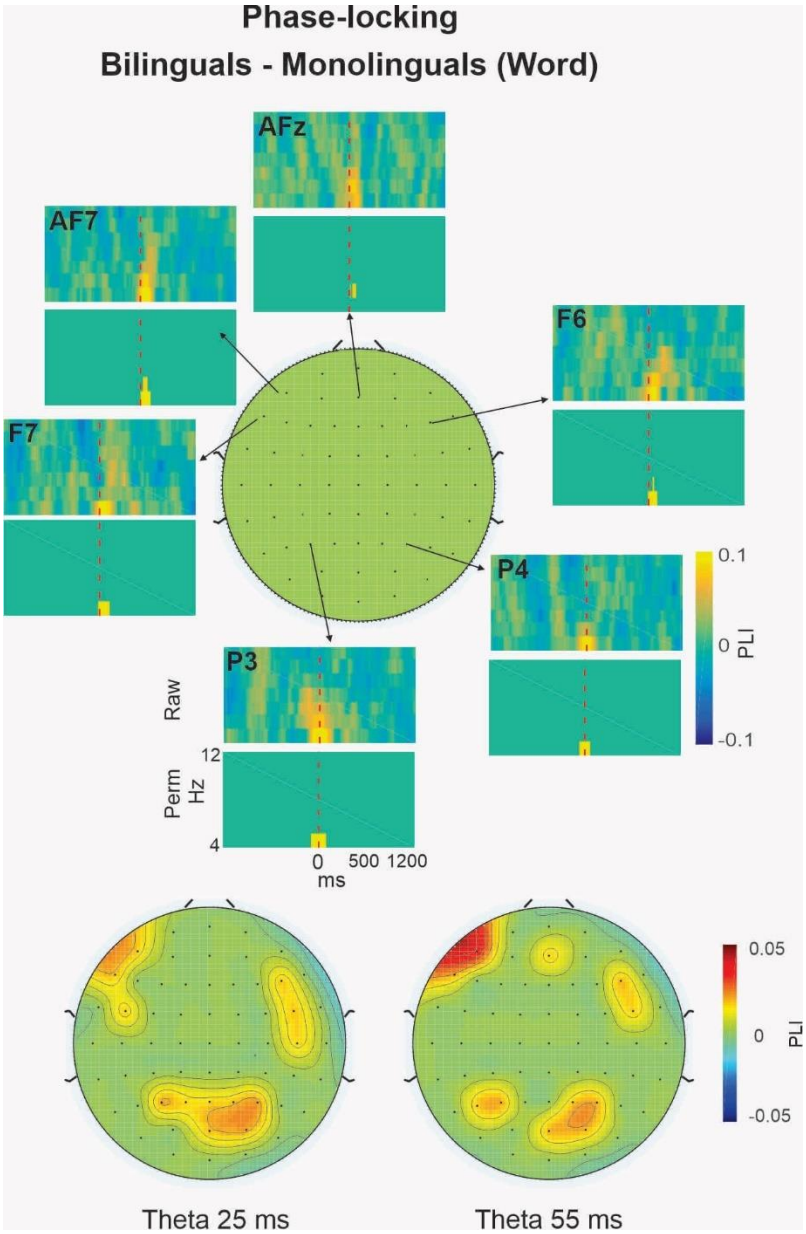

Figure 18: Phase-locking for Word Condition

Lastly, in the Pseudoword condition, there were no significant differences in phase-locking found between groups (no figures are presented).

4. Discussion

The phonemic restoration illusion is an auditory phenomenon where listeners perceptually fill in an interrupted speech signal with the missing phoneme (Warren, 1970). Prior research has shown that various top-down and bottom-up factors impact the degree to which listeners experience the illusion. However, most research on the

phonemic restoration effect has focused on native English speakers, with limited studies exploring how bilinguals experience this illusion (see Ishida and Arai (2016), who examined the perception of degraded speech in various bottom-up and top-down contexts among Japanese-English bilinguals). Furthermore, there is a notable gap concerning the underlying neural mechanisms of the illusion in bilinguals. Thus, this study's purpose was to better understand the effect of bilingualism on experiencing the phonemic restoration illusion and its underlying neural mechanisms.

We found significant differences between groups in both spectral power and phase-locking for each condition. Of importance is that for most conditions, bilinguals demonstrated an increase in theta response at the onset of the interruption for both spectral power and phase-locking. Only in the Passive condition did it differ in phase-locking results, with an increase in alpha around the onset. Alpha enhancement is associated with neural inhibition, suggesting perhaps that bilinguals are more likely to suppress auditory interruptions to focus on the visual silent movie during the Passive task (Jensen et al., 2002; Obleser & Weisz, 2012; Pfurtscheller et al., 1996). While prior research has indicated that theta power and phase-locking reduction reflects greater illusion perception (Riecke et al., 2009; Shahin et al., 2012), we cannot make these conclusions since there was no behavioral data collected during EEG recording. However, we did obtain behavioral results of illusion perception in the behavioral block. These results indicated a two-way interaction between lexical condition (Word or Pseudoword) and language group, such that monolinguals were more likely to fill-in pseudowords than words. Between language groups, there was no significant difference. In this case, we could expect to not see differences in theta power or phase-locking between groups for Word and Pseudoword conditions. However, looking at the spectral and phase-locking results for the Word condition, we observed greater theta phase-locking, enhanced alpha suppression, and gamma enhancement for bilinguals than monolinguals.

A possible explanation of this enhanced theta response could be that bilinguals have greater sensitivity to low-level acoustic cues, such as noise, resulting in stronger encoding of the interruption, compared to monolinguals. Because of this, they are more likely to heed attention to disruptions and noise which leads to stronger encoding of these bottom-up cues (Krizman et al., 2012; Skoe et al., 2017). Another possibility is that bilinguals are more likely to anticipate the onset of the interruption, which results in greater alpha suppression pre-onset and greater theta at the onset. Prior research has shown that anticipation results in enhanced alpha desynchronization (Foxe et al., 1998; Rohenkohl & Nobre, 2011; Simonet et al., 2019). From the "bilingual advantage" literature, there has been mixed results regarding bilingual performance in anticipatory tasks, with some research suggesting bilinguals outperform monolinguals in reaction time and accuracy on these tasks, and others finding that bilinguals perform worse than monolinguals (Amrhein, 1999; Bonifacci et al., 2011; Desideri & Bonifacci, 2018).

The pattern of greater alpha suppression followed by enhanced gamma after the onset of the interruption was found to be most distinct in the Word and /v/ conditions. This pattern has been found to indicate speech processing (Bauer et al., 2014; Nourski et al., 2022). Additionally, enhanced gamma has also been associated with the concept of

binding, the brain's ability to join certain features together (Tallon-Baudry & Bertrand, 1999; Tseng et al., 2016; White et al., 2018). Furthermore, the topographies for the Word and /v/ conditions suggest the involvement of motor regions, indicating that bilinguals engage motor areas more extensively than monolinguals. This aligns with the motor theory of speech perception, that speech perception is inherently connected to the mechanisms involved in speech production (Liberman et al., 1967; Liberman & Mattingly, 1985). According to this theory, hearing speech sounds triggers a mental production of the motor movements required to produce those sounds. Thus, in the present study, once a possible phoneme to fill-in the interruption has been selected, recruitment of motor regions occurs, followed by a binding of the phoneme with the rest of the word, which is reflected in the increased gamma activity.

We also see in the Pseudoword condition that while there is no enhanced alpha suppression, there is an increase in gamma power for bilinguals. This could be because pseudowords do not exist in the mental lexicon, resulting in less top-down recruitment compared to the Word condition. Perhaps due to this lack of being able to rely on top-down information, this results in a need for greater binding to occur, resulting in an increase in gamma. However, behaviorally, we did not find significant differences between groups in how pseudowords were perceived. Similar behavioral results were observed in Ishida and Arai (2016), where Japanese-English bilinguals exhibited no behavioral differences between lexical condition when restoring degraded speech.

Furthermore, in the Active condition, we observed greater alpha power suppression, followed by a more pronounced decrease in beta power occurring 500 ms after the noise onset for bilinguals. Desynchronization in the beta-band is often associated with recruitment of motor activity (Bartoli et al., 2016; Protzak & Gramann, 2021; Tzagarakis et al., 2010), once again aligning with the motor theory of speech perception, which suggests that motor regions are involved when processing speech (Liberman et al., 1967; Liberman & Mattingly, 1985). Within bilinguals, it has been suggested that more stress is placed upon the motor regions during speech perception, due to the complexities of learning to speak two languages (Simmonds et al., 2011). This might explain the greater decrease in beta activity observed among bilinguals, as these individuals may rely more heavily on motor involvement to accurately perceive speech.

Given these various findings, I suggest the following schematic to better understand the processes taking place between groups in various conditions (see Figure 19). This illustrates the differences in the processes that occur in bilinguals versus monolinguals when perceiving degraded speech, starting from the onset of the interruption at time 0 ms. The figure highlights differences in spectral power, as all conditions exhibited very similar phase-locking results of increased theta-phase locking at the noise onset (except for in the Pseudoword condition where there were no significant group differences in phase-locking).

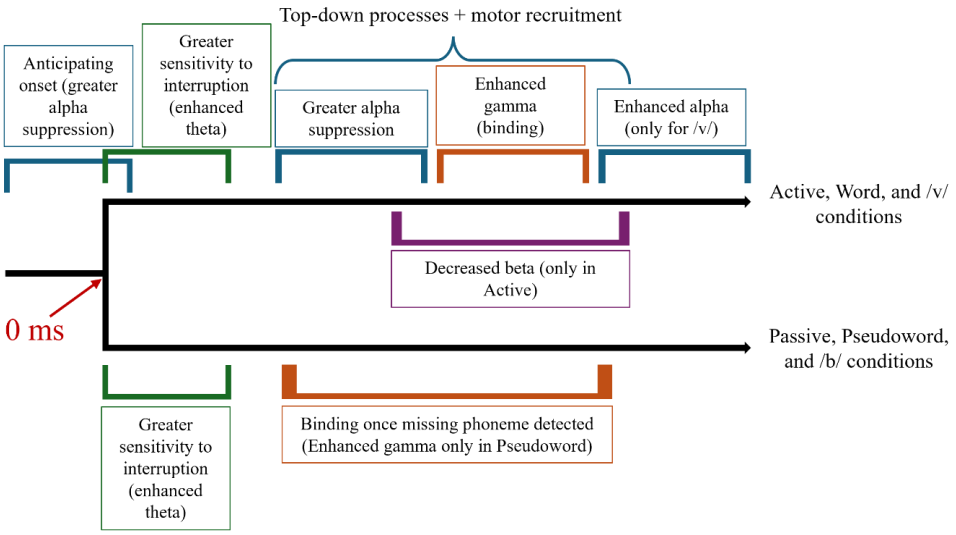

Figure 19: Schematic of proposed processes occurring in bilinguals when encountering degraded speech in various contexts. Descriptions describe bilingual processes relative to monolingual processes.

Starting before the interruption onset, in Active, Word, and /v/ conditions, bilinguals have greater alpha suppression, suggesting a degree of anticipation of perceiving the interruption. Given the complexity of these conditions—Active requires attention, Word involves lexical information, and /v/ is assumed to have a weaker phonetic weight for bilinguals—it is likely that bilinguals allocate more attention and engage more top-down processes for these conditions. This heightened cognitive demand could lead to greater expectation in these scenarios. Once encountering the interruption, there is an increase in theta response for bilinguals across all conditions, due to their greater sensitivity to bottom-up acoustic information. In the Word and /v/ conditions, bilinguals recruit top-down lexical networks to a greater extent than monolinguals, to fill in the gap after encountering the interruption. The complex phase-locking topographies in these conditions highlight this reliance on higher-level networks. This increased engagement of higher-level processes leads to greater alpha suppression. This is then followed by an increase in gamma power, suggesting that binding occurs between the missing phoneme and the target speech. In the Active condition, we instead see a decrease in beta-band activity, with no alpha suppression or gamma enhancement. As mentioned before, the results across these conditions could reflect stronger reliance upon top-down processes or engagement of motor processes. This suggests that stronger demands on placed on the motor and sensory regions in bilinguals due to learning how to articulate phonemes across two languages (Simmonds et al., 2011). Additionally, this decrease in beta-band activity could be attributed to bilinguals needing to inhibit one of their languages in certain contexts, a challenge monolinguals do not face. Prior research demonstrates that bilinguals activate both of their languages when listening to speech, and thus, they need to inhibit the non-target language to successfully select the intended speech in the target language (Kroll et al., 2008; Levy et al., 2007; Marian & Spivey, 2003; Spivey & Marian, 1999). Previous studies have demonstrated that bilinguals show beta reduction and increased alpha suppression during tasks involving code-switching,

suggesting that inhibiting one language may contribute to these neural changes (Geng et al., 2022; Timofeeva et al., 2023; Tomić & Kaan, 2022). In this study, since all the stimuli were in English and followed English phonotactics, it is possible that bilinguals needed to inhibit their Spanish during the conditions that required greater top-down recruitment.

For the Passive, Pseudoword, and /b/ conditions, we only noticed an increase in theta power at the onset of the interruption. Only in the Pseudoword condition was this followed by enhanced gamma, once again suggesting binding. Since pseudowords are not real words, there is less reliance on lexical knowledge. This explains the lack of enhanced alpha suppression, as higher-level processes are not engaged.

5. Conclusion

The current findings enhance our understanding of the mechanisms by which top-down and bottom-up factors modulate phonemic restoration in bilinguals. Although the results from this study are not entirely conclusive in explaining how bilinguals fill in degraded speech differently from monolingual listeners, the current results still shed light on the complex nature of bilingual speech perception. Further steps include conducting a behavioral perception task while collecting EEG data, incorporating both words and pseudowords in Spanish, and examining how language context shapes perception.

Conclusion

In this dissertation, I have analyzed how various top-down and bottom-up factors impact bilingual phonetic encoding from both behavioral and neurophysiological perspectives. I investigated how elements such as visual speech information, noise, attention and lexical knowledge influence bilinguals' phonemic encoding. This dissertation centered on Spanish-English bilinguals, primarily because of notable disparities in Spanish and English phonology as outlined earlier in the Introduction.

In Chapter 1, I investigated the mechanism of the McGurk illusion and the formation of "default" percepts in ambiguous scenarios, across words and pseudowords. By presenting auditory stimuli of words/pseudowords with missing phonemes, I found that when individuals fill-in a silent gap within a word/pseudoword with an incorrect phoneme, they fill-in with '/d/t/th/' more often than the other phonemes—especially for pseudowords. I found that when a viseme is ambiguous and not clearly discernible, this led to a failure in phonemic encoding by the visual cortex. As a result, the auditory system defaults to the naturally dominant and heavily weighted percept of '/d/t/th/.' Although this research did not specifically target bilinguals, it offered valuable insights into the processes of audiovisual speech perception, which is essential for discerning differences between bilinguals and monolinguals.

It is important to note, that Chapter 1 did not consider how prior knowledge and probabilistic models account for defaulting to perceiving 'd/t/th' in ambiguous situations. The Fuzzy Logical Model of Perception (FLMP) is a psycholinguistic theory that emphasizes that the brain uses probabilistic, rather than deterministic, processes to combine information from multiple sensory modalities (Massaro, 1987a, 1987b; Oden & Massaro, 1978). Under the FLMP, the information from the visual and auditory modalities are represented with varying degrees of certainty. Perception is determined by combining the information from both modalities. This involves calculating the degree of support for each possible phoneme from both modalities and then integrating these degrees to arrive at the most likely perceptual outcome. Future work should consider how FLMP and Bayesian models could possibly account for defaulting to certain phonemes under ambiguous speech contexts.

Chapter 2 delved deeper into audiovisual speech perception to explore its influence on the phonetic encoding of specific phonemes among Spanish-English bilinguals. Using a /va/ to /ba/-like continuum, I behaviorally measured how language experience modulates phonetic perception in auditory and audiovisual (AV) contexts, given that Spanish phonology does not distinguish between /b/ and /v/ unlike English. Participants were presented with AV congruent (e.g., auditory-ba paired with visual-ba), AV incongruent (e.g., auditory-ba paired with visual-va), and auditory-only stimuli and were asked to report what they heard. I found that in an auditory-only setting, lower English dominance resulted in greater perception of 'ba,' only when a /va/ was played. Additionally, for 'va' perception, weaker English dominance resulting in less 'va' perception, when a /va/ was presented. However, as the continuum shifted more towards /ba/, we found that less English dominance resulted in an increase in 'va' perception, suggesting that under ambiguous phonetic situations, individuals with lower English

dominance shift their phonetic boundaries to adjust for an English context. In audiovisual contexts, bilinguals were more susceptible to visual influence, especially when the auditory stimulus was a /va/. This study highlighted the complex relationship between language dominance and phonetic perception in bilingual individuals.

Finally, in Chapter 3, I investigated the neural mechanisms that underlie the phonemic restoration effect—a phenomenon where listeners can perceptually fill in degraded speech signals despite extraneous interruptions (Warren, 1970)—between Spanish-English bilinguals and English monolinguals. I also explored the influence of top-down effects such as attention and lexical condition (word/pseudoword), as well as bottom-up acoustic factors of the phoneme omitted, on the neural mechanisms involved in the restoration of degraded speech, along with the perceptual ability of restoring speech. I found that at the onset of an interruption in speech, bilinguals exhibited increased theta power and phase-locking across lexical condition, whether a /b/ or /v/ phoneme was removed, and when attention was employed, suggesting greater sensitivity to bottom-up acoustic information for bilinguals. Additionally, this was followed with bilinguals showing greater alpha suppression and enhanced gamma in Active, Word, and /v/ conditions. Given the complexity of these conditions—Active requires attention, Word involves lexical information, and /v/ is assumed to have a weaker phonetic weight for bilinguals—this indicates that compared to monolinguals, bilinguals potentially rely more heavily on top-down lexical knowledge and show an increase in binding between the missing phoneme and the target speech.

Overall, this dissertation contributes to our understanding of how bilingualism impacts speech perception and how bottom-up and top-down factors interact within bilinguals compared to monolinguals. These experiments have shed light on how the unique experience of being bilingual impacts speech perception, examining both the behavioral and neural manifestations in bilingual individuals. Through a series of behavioral and neurophysiological studies, it becomes evident that bilinguals' phonemic perception is intricately influenced by language experience, visual context, top-down factors, and phonological inventory.

Additionally, the experiments in this dissertation contribute to our understanding of the mechanisms behind speech perception in ambiguous situations. When we encounter degraded speech, our brain initially tries to fill in the missing speech sounds, as demonstrated by the phenomenon of phonemic restoration. This suggests that phonemic restoration is a primary repair mechanism, occurring at a low-level auditory processing stage, since similar processes have been observed in non-speech situations. If the brain cannot adequately fill in the missing phoneme using phonemic restoration, it may default to substituting sounds like 'd/t/th,' as observed in Chapter 1. This indicates a secondary repair strategy, where the brain attempts to compensate for the unclear input by using common phonetic patterns. In cases where neither phonemic restoration nor defaulting occurs, the degraded speech is perceived as it is, with the gap being encoded and the missing phoneme left unfilled. This sequence of strategies highlights the brain's adaptability in handling incomplete auditory information.

A key question that remains in the field of bilingual research is determining the best measure to use when studying bilinguals. There is considerable debate over the most effective method to identify bilingual differences, with factors such as age of acquisition, literacy, language dominance, vocabulary size, and more being considered (Backer & Bortfeld, 2021). In my dissertation, I used a scale of language dominance to assess participants in Chapter 2, while in Chapter 3, I categorized participants into bilingual or monolingual groups. Although my dissertation does not provide a definitive answer to the best measure, it is crucial to acknowledge that bilinguals are not a homogeneous group. To better understand the effects of bilingualism, we must recognize the individual differences across various language factors among bilinguals. Ideally, my dissertation would have included a broader range of measures, such as literacy tests in both languages, vocabulary size assessments, and an understanding of language use, among others.

As the population of bilinguals continues to grow in the United States, understanding the distinctive aspects of bilingual language processing becomes increasingly important. Such insights are crucial for evaluating the potential of bilingual immersion programs in American education and their implications. Educators focusing on heritage and bilingual language pedagogy benefit from bilingual research, particularly given the expanding bilingual population in the US. As this population grows, comprehending the effects of bilingualism becomes imperative for adapting educational practices to effectively accommodate linguistic diversity.

References

Abbott, N. T., & Shahin, A. J. (2018). Cross-modal phonetic encoding facilitates the McGurk illusion and phonemic restoration. *Journal of Neurophysiology*, *120*(6), 2988–3000. https://doi.org/10.1152/jn.00262.2018

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, *1*(1), 1–8. https://doi.org/10.1016/S0095-4470(19)31372-5

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689–698. https://doi.org/10.1017/S1366728916000225

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty Years After Hearing Lips and Seeing Voices: The McGurk Effect Revisited. *Multisensory Research*, *31*(1–2), 111–144. https://doi.org/10.1163/22134808-00002565

Amrhein, P. C. (1999). On the Functional Equivalence of Monolinguals and Bilinguals in "Monolingual Mode": The Bilingual Anticipation Effect in Picture-Word Processing. *Psychological Science*, *10*(3), 230–236. https://doi.org/10.1111/1467-9280.00142

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

Ansaldo, A. I., Marcotte, K., Scherer, L., & Raboyeau, G. (2008). Language therapy and bilingual aphasia: Clinical implications of psycholinguistic and neuroimaging research. *Journal of Neurolinguistics*, *21*(6), 539–557. https://doi.org/10.1016/j.jneuroling.2008.02.001

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual Neural Routing of Visual Facilitation in Speech Processing. *Journal of Neuroscience*, *29*(43), 13445–13453. https://doi.org/10.1523/JNEUROSCI.3194-09.2009

Backer, K. C., & Bortfeld, H. (2021). Characterizing Bilingual Effects on Cognition: The Search for Meaningful Individual Differences. *Brain Sciences*, *11*(1), Article 1. https://doi.org/10.3390/brainsci11010081

Barragan, B., Liss, J., Berisha, V., Santello, M., & Uehara, K. (2022). *Shared neural substrates for second-language speech production and word recognition*. College of Health Solutions, Arizona State University.

Bartoli, E., Maffongelli, L., Campus, C., & D'Ausilio, A. (2016). Beta rhythm modulation by speech sounds: Somatotopic mapping in somatosensory cortex. *Scientific Reports*, *6*(1), 31182. https://doi.org/10.1038/srep31182

Bashford, J. A., Warren, R. M., & Brown, C. A. (1996). Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. *Perception & Psychophysics*, *58*(3), 342–350. https://doi.org/10.3758/BF03206810

Basu Mallick, D., F. Magnotti, J., & S. Beauchamp, M. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, M., Stenner, M.-P., Friston, K. J., & Dolan, R. J. (2014). Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes. *Journal of Neuroscience*, *34*(48), 16117–16125. https://doi.org/10.1523/JNEUROSCI.3474-13.2014

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron*, *41*(5), 809–823. https://doi.org/10.1016/S0896-6273(04)00070-4

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect. *Journal of Neuroscience*, *30*(7), 2414–2417. https://doi.org/10.1523/JNEUROSCI.4865-09.2010

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(52), 14301–14310. https://doi.org/10.1523/JNEUROSCI.2875-08.2008

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Best, C. T. (1994). *The emergence of native-language phonological influences in infants: A perceptual assimilation model. The development of speech perception: The transition from speech sounds to spoken words, ed. By Judith Goodman and Howard Nusbaum, 167–224*. Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. https://doi.org/10.1121/1.1332378

Bialystok, E. (2010). Bilingualism. *WIREs Cognitive Science*, *1*(4), 559–572. https://doi.org/10.1002/wcs.43

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Binder, J., & Price, C. (2000). Functional neuroimaging of language process. *Handbook of Functional Imaging of Cognition, Ed. R. Cabeza & A. Kingstone. MIT Press.[CW]*.

Blank, H., & von Kriegstein, K. (2013). Mechanisms of enhancing visual–speech recognition by prior auditory information. *NeuroImage*, *65*, 109–118. https://doi.org/10.1016/j.neuroimage.2012.09.047

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, *14*(2), 256–269. https://doi.org/10.1111/j.1467-7687.2010.00974.x

Borghini, G., & Hazan, V. (2018). Listening Effort During Sentence Processing Is Increased for Non-native Listeners: A Pupillometry Study. *Frontiers in Neuroscience*, *12*, 152. https://doi.org/10.3389/fnins.2018.00152

Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect? *PLOS ONE*, *13*(11), e0207160. https://doi.org/10.1371/journal.pone.0207160

Bsharat-Maalouf, D., & Karawani, H. (2022). Bilinguals' speech perception in noise: Perceptual and neural associations. *PLOS ONE*, *17*(2), e0264282. https://doi.org/10.1371/journal.pone.0264282

Burfin, S., Pascalis, O., Ruiz Tada, E., Costa, A., Savariaux, C., & Kandel, S. (2014). Bilingualism affects audiovisual phoneme identification. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01179

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657. https://doi.org/10.1016/S0960-9822(00)00513-3

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428. https://doi.org/10.1121/1.1913594

Casillas, J. V., & Simonet, M. (2018). *Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.wocn.2018.07.002

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. https://doi.org/10.1038/nn.2641

Chauvin, A., Pellerin, S., Boatswain-Jacques, A.-F., René, J.-L., & Phillips, N. A. (2024). Audiovisual speech perception in noise in younger and older bilinguals. *Psychology and Aging*, *39*(3), 245–261. https://doi.org/10.1037/pag0000799

Chauvin, A., & Phillips, N. A. (2022). Bilinguals Show Proportionally Greater Benefit From Visual Speech Cues and Sentence Context in Their Second Compared to Their First Language. *Ear and Hearing*, *43*(4), 1316–1326. https://doi.org/10.1097/AUD.0000000000001182

Chen, Y., & Hazan, V. (2007). Language effects on the degree of visual influence in audiovisual speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrueken, Germany*, 6–10.

Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife*, *5*, e12577. https://doi.org/10.7554/eLife.12577

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, *19*(5), 381–385. https://doi.org/10.1016/j.cub.2009.01.017

Desideri, L., & Bonifacci, P. (2018). Verbal and Nonverbal Anticipatory Mechanisms in Bilinguals. *Journal of Psycholinguistic Research*, *47*(3), 719–739. https://doi.org/10.1007/s10936-017-9556-1

Elff, M. (2022). *mclogit: Multinomial Logit Models, with or without Random Effects or Overdispersion* [Computer software].

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *J. Acoust. Soc. Am.*, *62*(4), 5.

Erickson, L. C., Zielinski, B. A., Zielinski, J. E. V., Liu, G., Turkeltaub, P. E., Leaver, A. M., & Rauschecker, J. P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00534

Feng, G., Zhou, B., Zhou, W., Beauchamp, M. S., & Magnotti, J. F. (2019). A Laboratory Study of the McGurk Effect in 324 Monozygotic and Dizygotic Twins. *Frontiers in Neuroscience*, *13*, 1029. https://doi.org/10.3389/fnins.2019.01029

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

Florentine, M. (1985). Speech perception in noise by fluent, non-native listeners. *The Journal of the Acoustical Society of America*, *77*(S1), S106–S106. https://doi.org/10.1121/1.2022152

Foxe, J. J., Simpson, G. V., & Ahlfors, S. P. (1998). Parieto-occipital ~1 0Hz activity reflects anticipatory state of visual attention mechanisms. *NeuroReport*, *9*(17), 3929.

Fricke, M. (2022). Modulation of Cross-Language Activation During Bilingual Auditory Word Recognition: Effects of Language Experience but Not Competing Background Noise. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.674157

Fricke, M., & Zirnstein, M. (2022). Predictive Processing and Inhibitory Control Drive Semantic Enhancements for Non-Dominant Language Word Recognition in Noise. *Languages*, *7*(3), Article 3. https://doi.org/10.3390/languages7030239

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. https://doi.org/10.3758/BF03193857

Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*(4), 369–378. https://doi.org/10.1016/j.specom.2008.11.005

García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, *121*(3), 194–205. https://doi.org/10.1016/j.bandl.2012.03.008

Geng, S., Molinaro, N., Timofeeva, P., Quiñones, I., Carreiras, M., & Amoruso, L. (2022). Oscillatory dynamics underlying noun and verb production in highly proficient bilinguals. *Scientific Reports*, *12*(1), 764. https://doi.org/10.1038/s41598-021-04737-z

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging Studies of Language Production and Comprehension. *Annual Review of Psychology*, *54*, 91–114. https://doi.org/10.1146/annurev.psych.54.101601.145128

Getz, L. M., & Toscano, J. C. (2021). Rethinking the McGurk effect as a perceptual illusion. *Attention, Perception, & Psychophysics*, *83*(6), 2583–2598. https://doi.org/10.3758/s13414-021-02265-6

Ghazanfar, A. A. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience*, *25*(20), 5004–5012. https://doi.org/10.1523/JNEUROSCI.0799-05.2005

Gonzales, M. G., Backer, K. C., Mandujano, B., & Shahin, A. J. (2021). Rethinking the Mechanisms Underlying the McGurk Illusion. *Frontiers in Human Neuroscience*, *15*, 616049. https://doi.org/10.3389/fnhum.2021.616049

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197. https://doi.org/10.1121/1.1288668

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research*, *1361*, 54–66. https://doi.org/10.1016/j.brainres.2010.09.003

Grosjean, F. (1994). *PII: 0093-934X(85)90048-3 | Elsevier Enhanced Reader*. https://doi.org/10.1016/0093-934X(85)90048-3

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333–1341. https://doi.org/10.3758/s13414-014-0821-1

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, *36*(1), 17–38.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Hirst, R. J., Stacey, J. E., Cragg, L., Stacey, P. C., & Allen, H. A. (2018). The threshold for the McGurk effect in audio-visual noise decreases with development. *Scientific Reports*, *8*(1), 12372. https://doi.org/10.1038/s41598-018-30798-8

Hocking, J., & Price, C. J. (2008). The Role of the Posterior Superior Temporal Sulcus in Audiovisual Processing. *Cerebral Cortex*, *18*(10), 2439–2449. https://doi.org/10.1093/cercor/bhn007

Hwang, J., & Romanski, L. M. (2015). Prefrontal Neuronal Responses during Audiovisual Mnemonic Processing. *Journal of Neuroscience*, *35*(3), 960–971. https://doi.org/10.1523/JNEUROSCI.1328-14.2015

Ingvalson, E. M., Holt, L. L., & McCLELLAND, J. L. (2012). Can native Japanese listeners learn to differentiate /r–l/ on the basis of F3 onset frequency? – CORRIGENDUM. *Bilingualism: Language and Cognition*, *15*(2), 434–435. https://doi.org/10.1017/S1366728912000041

Ishida, M., & Arai, T. (2016). Missing phonemes are perceptually restored but differently by native and non-native listeners. *SpringerPlus*, *5*(1), 713. https://doi.org/10.1186/s40064-016-2479-8

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the Alpha Band (9–12 Hz) Increase with Memory Load during Retention in a Short-term Memory Task. *Cerebral Cortex*, *12*(8), 877–882. https://doi.org/10.1093/cercor/12.8.877

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual Enhancement of the Information Representation in Auditory Cortex. *Current Biology*, *20*(1), 19–24. https://doi.org/10.1016/j.cub.2009.10.068

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex*, *18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(2), 634–647. https://doi.org/10.1037//0096-1523.26.2.634

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). *Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages*. https://www.pnas.org/doi/epdf/10.1073/pnas.1201575109

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two Languages in Mind: Bilingualism as a Tool to Investigate Language, Cognition, and the Brain. *Current Directions in Psychological Science*, *23*(3), 159–163. https://doi.org/10.1177/0963721414528511

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, *255*(5044), 606–608.

Lecumberri, M., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Levi, S. V. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition*, *21*(3), 523–536. https://doi.org/10.1017/S1366728917000153

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, *18*(1), 29–34.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.
Lisker, L., & Abramson, A. S. (1970). *The voicing dimension: Some experiments in comparative phonetics*. 563–567.
MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390. https://doi.org/10.1017/S0142716400009796
Marian, V., Hayakawa, S., Lam, T. Q., & Schroeder, S. R. (2018). Language Experience Changes Audiovisual Perception. *Brain Sciences*, *8*(5), Article 5. https://doi.org/10.3390/brainsci8050085
Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(2), 173–193. https://doi.org/10.1017/S0142716403000092
Massaro, D. W. (1987a). Categorical partition: A fuzzy-logical model of categorization behavior. In *Categorical perception: The groundwork of cognition.* (pp. 254–283). Cambridge University Press.
Massaro, D. W. (1987b). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Lawrence Erlbaum Associates, Publishers.
Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693. https://doi.org/10.1044/jslhr.4003.686
McDonald, M., & Kaushanskaya, M. (2023). Bilingual Children Shift and Relax Second-Language Phoneme Categorization in Response to Accented L2 and Native L1 Speech Exposure. *Language and Speech*, 00238309231176760. https://doi.org/10.1177/00238309231176760
McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0
Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science (New York, N.Y.)*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994
Miller, B. T. & D'Esposito. (2005). Searching for "the top" in top-down control. *Neuron*, *48*, 535–538. https://doi.org/10.1111/j.1530-0277.2011.01535.x
Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352. https://doi.org/10.1121/1.1907526
Miller, L. M., & D'Esposito, M. (2005). Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech. *Journal of Neuroscience*, *25*(25), 5884–5893. https://doi.org/10.1523/JNEUROSCI.0896-05.2005
Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340. https://doi.org/10.3758/BF03211209
Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, 283–303.
Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J. P., Schwartz, T. H., & Foxe, J. J. (2006). Audio-Visual Multisensory

Integration in Superior Parietal Lobule Revealed by Human Intracranial Recordings. *Journal of Neurophysiology*, *96*(2), 721–729. https://doi.org/10.1152/jn.00285.2006

Montrul, S. (2011). INTRODUCTION: The Linguistic Competence of Heritage Speakers. *Studies in Second Language Acquisition*, *33*(2), 155–161. https://doi.org/10.1017/S0272263110000719

Morís Fernández, L., Macaluso, E., & Soto-Faraco, S. (2017). Audiovisual integration as conflict resolution: The conflict of the McGurk illusion: The Conflict of the McGurk Illusion. *Human Brain Mapping*, *38*(11), 5691–5705. https://doi.org/10.1002/hbm.23758

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12. https://doi.org/10.1007/s00426-005-0031-5

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual Decisions Formed by Accumulation of Audiovisual Evidence in Prefrontal Cortex. *Journal of Neuroscience*, *30*(21), 7434–7446. https://doi.org/10.1523/JNEUROSCI.0455-10.2010

Nourski, K. V., Steinschneider, M., Rhone, A. E., Kovach, C. K., Kawasaki, H., & Howard, M. A. (2022). Gamma Activation and Alpha Suppression within Human Auditory Cortex during a Speech Classification Task. *Journal of Neuroscience*, *42*(25), 5034–5046. https://doi.org/10.1523/JNEUROSCI.2187-21.2022

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(11), 2466–2477. https://doi.org/10.1093/cercor/bhr325

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172.

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., Weizman, A., & Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Comprehensive Psychiatry*, *50*(2), 186–192. https://doi.org/10.1016/j.comppsych.2008.06.004

Pfurtscheller, G., Stancák, A., & Neuper, Ch. (1996). Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, *24*(1), 39–46. https://doi.org/10.1016/S0167-8760(96)00066-9

Pilling, M. (2009). Auditory Event-Related Potentials (ERPs) in Audiovisual Speech Perception. *Journal of Speech, Language, and Hearing Research*, *52*(4), 1073–1081. https://doi.org/10.1044/1092-4388(2009/07-0276)

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d /–/ð / perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190–2201. https://doi.org/10.1121/1.1362689

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, *26*(4), 490–498. https://doi.org/10.1177/0956797614568320

Protzak, J., & Gramann, K. (2021). EEG beta-modulations reflect age-specific motor resource allocation during dual-task walking. *Scientific Reports*, *11*(1), 16110. https://doi.org/10.1038/s41598-021-94874-2

Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific Reports*, *6*(1), 30423. https://doi.org/10.1038/srep30423

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Riecke, L., Esposito, F., Bonte, M., & Formisano, E. (2009). Hearing Illusory Sounds in Noise: The Timing of Sensory-Perceptual Transformations in Auditory Cortex. *Neuron*, *64*(4), 550–561. https://doi.org/10.1016/j.neuron.2009.10.016

Rogers, B. (2022). When is an illusion not an illusion? An alternative view of the illusion concept. *Frontiers in Human Neuroscience*, *16*, 957740. https://doi.org/10.3389/fnhum.2022.957740

Rohenkohl, G., & Nobre, A. C. (2011). Alpha Oscillations Related to Anticipatory Attention Follow Temporal Expectations. *Journal of Neuroscience*, *31*(40), 14076–14084. https://doi.org/10.1523/JNEUROSCI.3387-11.2011

Romanski, L. M. (2012). *Convergence of auditory, visual, and somatosensory information in ventral prefrontal cortex*.

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception & Psychophysics*, *52*(4), 461–473. https://doi.org/10.3758/BF03206706

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–494. https://doi.org/10.1037/0096-3445.110.4.474

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1124–1131. https://doi.org/10.1037/0096-1523.7.5.1124

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. https://doi.org/10.3758/BF03202653

Samuel, A. G. (1997). Lexical Activation Produces Potent Phonemic Percepts. *Cognitive Psychology*, *32*(2), 97–127. https://doi.org/10.1006/cogp.1997.0646

Samuel, A. G., & Newport, E. L. (1979). Adaptation of Speech by Nonspeech: Evidence for Complex Acoustic Cue Detectors. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(3), 563–578. https://doi.org/10.1037/h0078136

Schmitz, J., Bartoli, E., Maffongelli, L., Fadiga, L., Sebastian-Galles, N., & D'Ausilio, A. (2019). Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia*, *128*, 290–296. https://doi.org/10.1016/j.neuropsychologia.2018.01.006

Schmitz, J., Díaz, B., Fernández Rubio, K., & Sebastian-Galles, N. (2018). Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Language, Cognition and Neuroscience*, *33*(5), 527–546. https://doi.org/10.1080/23273798.2017.1390142

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the*

National Academy of Sciences, *102*(51), 18748–18750. https://doi.org/10.1073/pnas.0508862102

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400

Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan (E)*, *15*(3), 143–158. https://doi.org/10.1250/ast.15.143

Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00323

Shahin, A. J., Backer, K. C., Rosenblum, L. D., & Kerlin, J. R. (2018). Neural Mechanisms Underlying Cross-Modal Phonetic Encoding. *The Journal of Neuroscience*, *38*(7), 1835–1849. https://doi.org/10.1523/JNEUROSCI.1566-17.2017

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage*, *44*(3), 1133–1143. https://doi.org/10.1016/j.neuroimage.2008.09.045

Shahin, A. J., Kerlin, J. R., Bhat, J., & Miller, L. M. (2012). Neural restoration of degraded audiovisual speech. *NeuroImage*, *60*(1), 530–538. https://doi.org/10.1016/j.neuroimage.2011.11.097

Shahin, A. J., & Miller, L. M. (2009). Multisensory integration enhances phonemic restoration. *The Journal of the Acoustical Society of America*, *125*(3), 1744–1750. https://doi.org/10.1121/1.3075576

Shatzer, H., Shen, S., Kerlin, J. R., Pitt, M. A., & Shahin, A. J. (2018). Neurophysiology underlying influence of stimulus reliability on audiovisual integration. *European Journal of Neuroscience*, *48*(8), 2836–2848. https://doi.org/10.1111/ejn.13843

Sherman, G. (1971). The phonemic restoration effect: An insight into the mechanisms of speech perception. *Unpublished Master's Thesis, University of WisconsinMilwaukee*.

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two Tongues, One Brain: Imaging Bilingual Speech Production. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00166

Simonet, M., Meziane, H. B., Runswick, O. R., North, J. S., Williams, A. M., Barral, J., & Roca, A. (2019). The modulation of event-related alpha rhythm during the time course of anticipation. *Scientific Reports*, *9*(1), 18226. https://doi.org/10.1038/s41598-019-54763-1

Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: Evidence from early and late ERP effects. *Brain Research*, *1121*(1), 177–189. https://doi.org/10.1016/j.brainres.2006.08.123

Skoe, E., Burakiewicz, E., Figueiredo, M., & Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience*, *349*, 278–290. https://doi.org/10.1016/j.neuroscience.2017.02.049

Skoe, E., & Karayanidi, K. (2019). Bilingualism and Speech Understanding in Noise: Auditory and Linguistic Factors. *Journal of the American Academy of Audiology*, *30*(02), 115–130. https://doi.org/10.3766/jaaa.17082

Smith, E., Duede, S., Hanrahan, S., Davis, T., House, P., & Greger, B. (2013). Seeing Is Believing: Neural Representations of Visual Stimuli in Human Auditory Cortex Correlate with Illusory Auditory Perceptions. *PLoS ONE*, *8*(9), e73148. https://doi.org/10.1371/journal.pone.0073148

Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Stekelenburg, J. J., & Vroomen, J. (2007). Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

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

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America*, *88*(2), 663–666. https://doi.org/10.1121/1.399769

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Timofeeva, P., Quiñones, I., Geng, S., de Bruin, A., Carreiras, M., & Amoruso, L. (2023). Behavioral and oscillatory signatures of switch costs in highly proficient bilinguals. *Scientific Reports*, *13*(1), 7725. https://doi.org/10.1038/s41598-023-34895-1

Tomić, A., & Kaan, E. (2022). Oscillatory brain responses to processing code-switches in the presence of others. *Brain and Language*, *231*, 105139. https://doi.org/10.1016/j.bandl.2022.105139

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and Non-Speech Audio-Visual Illusions: A Developmental Study. *PLoS ONE*, *2*(8), e742. https://doi.org/10.1371/journal.pone.0000742

Tseng, P., Chang, Y.-T., Chang, C.-F., Liang, W.-K., & Juan, C.-H. (2016). The critical role of phase difference in gamma oscillation within the temporoparietal network for binding visual working memory. *Scientific Reports*, *6*(1), 32138. https://doi.org/10.1038/srep32138

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-Band Activity during Motor Planning Reflects Response Uncertainty. *The Journal of Neuroscience*, *30*(34), 11270–11277. https://doi.org/10.1523/JNEUROSCI.6026-09.2010

Valdés, G., & Anderson, N. (2000). *American Association of Teachers of Spanish and Portuguese professional development series handbook for teachers K–16: Vol. I. Spanish for native speakers*. San Diego, CA: Harcourt College Publishers.

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. *The Journal of the Acoustical Society of America*, *152*(6), 3216–3225. https://doi.org/10.1121/10.0015262

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, *79*(2), 396–403. https://doi.org/10.3758/s13414-016-1238-9

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00174

Warren, R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, *167*(3917), 392–393. https://doi.org/10.1126/science.167.3917.392

Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., & Healy, E. W. (1997). Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps. *Perception & Psychophysics*, *59*(2), 275–283. https://doi.org/10.3758/BF03211895

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, *9*(3), 358–362. https://doi.org/10.3758/BF03212667

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development*, *52*(1), 349–355. https://doi.org/10.2307/1129249

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, E. J., Nayman, C., Dunkley, B. T., Keller, A. E., Valiante, T. A., & Pang, E. W. (2018). Addressing the Language Binding Problem With Dynamic Functional Connectivity During Meaningful Spoken Language Comprehension. *Frontiers in Psychology*, *9*, 1960. https://doi.org/10.3389/fpsyg.2018.01960

Wig, N., & García-Sierra, A. (2021). Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals' speech perception. *Bilingualism: Language and Cognition*, *24*(3), 467–480. https://doi.org/10.1017/S1366728920000553

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289–297. https://doi.org/10.3758/BF03199477

Wilson, S., Marks, R., Collins, N., Warner, B., & Frick, L. (2004). Benefits of multidisciplinary case conferencing using audiovisual compared with telephone communication: A randomized controlled trial. *Journal of Telemedicine and Telecare*, *10*(6), 351–354.

Zhu, L. L., & Beauchamp, M. S. (2017). Mouth and Voice: A Relationship between Visual and Auditory Preference in the Human Superior Temporal Sulcus. *The Journal of Neuroscience*, *37*(10), 2697–2708. https://doi.org/10.1523/JNEUROSCI.2914-16.2017