Individual Differences in Speech Perception: The Malleability of Gradiency and the Functions of Gradiency and Consistency

Wai Leung Wong

Supervised by

Dr. Efthymia Kapnoula & Dr. Arthur Samuel



2026

Individual Differences in Speech Perception: The Malleability of Gradiency and the Functions of Gradiency and Consistency

Doctoral dissertation by
Wai Leung Wong

Dissertation submitted for the degree of
Doctor of Philosophy

Supervised by
Dr. Efthymia Kapnoula & Dr. Arthur Samuel



Universidad del País Vasco

Euskal Herriko Unibertsitatea



BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

Donostia-San Sebastián, 2026

This research was supported by a predoctoral fellowship (grant PRE2021-097223, associated with project CEX2020-001010-S–21–2) funded by the Spanish Ministry of Science and Innovation (MCIN), the State Research Agency (AEI), and the European Social Fund Plus (FSE+).

Wai Leung Wong
All rights reserved.

Basque Center on Cognition, Brain and Language
Paseo Mikeletegi, 69,
Donostia-San Sebastián, Spain
March 2026

Acknowledgements

I am deeply grateful to my supervisors, Efthymia Kapnoula and Arthur Samuel, for their continuous guidance, support, and encouragement throughout my PhD. They introduced me to the world of speech perception at a time when I was only beginning to explore it, long before I fully understood how fortunate I was to have them as mentors. Effie, I deeply appreciate your kindness and the time you devoted to meeting with me every week. Arty, I am thankful for your exceptional responsiveness whenever I needed advice, and for guiding me through many aspects of academic life, from writing papers to navigating the broader research community. Thank you both for showing me what thoughtful and supportive supervision truly looks like, and for trusting me as a collaborator rather than only a student.

Beyond my supervisors, I would like to thank all other current and former members of the Spoken Language Interest Group (SLIG) for creating a stimulating and supportive environment for sharing ideas. I am especially thankful to Drew McLaughlin for her mentorship and for teaching me pupillometry and growth curve analysis.

I would also like to thank the current and former lab members at BCBL. Huge thanks go to Itziar Basterra; without her incredible effort in recruiting and scheduling participants, these PhD projects could not have been completed. I am also grateful to Ainhoa Eguiguren, Amets Esnal, Manex Lete, Elena Aguirrebengoa, Lidia Insausti, Olatz Andonegui, and Janire Sarasua for their help with laboratory logistics, and to Larraitz Lazcano, Eider Juaristi, and Maider Goni for patiently answering my questions about the PhD process.

I am very glad to have spent about four months at the Language and Brain (LAB) Lab at the University of Connecticut for my research stay. I am thankful to Emily Myers for hosting me and for sharing many stimulating ideas. I would also like to thank all members of the LAB Lab for their hospitality and for making my stay intellectually enriching, especially Portia Washington, for her generous help and constant support during my time there; Hannah

Olson, for organizing data collection for Experiment 1.2 in this dissertation; and the undergraduate research assistants who helped with data collection. Thanks also go to Shawn Cummings, Anne Marie Crinnion, Holly Zaharchuk, and Hannah Mechtenberg for their constructive comments on my projects.

I am grateful to Brendan Costello and James Magnuson for serving on my midterm committee and for providing constructive and insightful feedback on the project presented in Chapter 3. I would also like to thank all colleagues at BCBL who provided constructive feedback on my projects throughout these years.

Special thanks to my friends Silvia Yang, Irene Qiu, and Yiming Lu at BCBL, whose kindness and companionship made my years there brighter and more enjoyable. I am especially grateful to Irene Qiu for her extensive expertise in phonetics and Praat, which was invaluable to multiple experiments in this dissertation. I also thank Yiming Lu for his contributions to the computational modeling related to Experiment 1.

I am also grateful to Ethan Kutlu at the University of Iowa for his generous support and for providing me with the opportunity and funding to pursue a project I am passionate about. I also thank Hyoju Kim for insightful discussions and feedback that helped shape some of the ideas in this dissertation.

I would like to acknowledge all my mentors and collaborators at the Chinese University of Hong Kong, especially Urs Maurer and Tomohiro Inoue, for their continued guidance and support even during my PhD.

Above all, I am profoundly grateful to my family and friends in Hong Kong for their unconditional love, encouragement, and understanding. Their unwavering support has sustained me through every challenge of this journey.

Abstract

This dissertation investigates the multidimensional nature of subphonemic speech perception, focusing on two constructs: gradiency, defined as sensitivity to within-phoneme acoustic differences, and consistency, defined as the stability of responses to the same speech sound across repetitions. While previous research has shown that listeners are generally gradient, substantial individual differences exist. Similar individual differences are also observed in consistency. However, the sources of these individual differences, the malleability of gradiency and consistency, and the functional consequences of these constructs remain largely unclear.

To this end, across three experiments, this dissertation examines the malleability of gradiency and the functional roles of gradiency and consistency, with particular emphasis on bilabial voicing contrasts (/b/–/p/) in Spanish, Basque, and English, using voice onset time (VOT) as the primary acoustic cue. These experiments also shed light on the sources and characteristics of both constructs. By focusing on multilingual listeners and incorporating multiple languages, this dissertation extends a literature that has been largely based on monolingual English listeners.

Experiment 1 focuses on the malleability of gradiency using a distributional learning paradigm. It includes two sub-experiments that test whether bilingual listeners adjust their perception of the /b/–/p/ contrast based on the variance of VOT distributions. Seventy-two Spanish–Basque bilinguals (Experiment 1.1) and 72 Spanish–English bilinguals (Experiment 1.2) were exposed to narrow or wide VOT distributions in each language, operationalized through the distribution of trial frequencies across VOT steps, with either the same or mixed variance across languages. Distributional learning was assessed using two measures: the slope of categorization functions based on mouse-click responses and the proportion of

fixations to phonological competitors in a Visual World Paradigm task. Previous studies with monolingual English listeners have shown that wider distributions yield shallower categorization slopes and increased competitor fixations relative to narrower distributions, indicating sensitivity to distributional variance. In the present study, however, this pattern was not observed consistently. Exposure to wider VOT distributions produced shallower categorization slopes and increased competitor fixations in some conditions, but the strength of these effects varied depending on the language pair and the measure used to index speech perception. Specifically, Spanish–Basque bilinguals showed increased competitor fixations in the wide compared to the narrow condition, but no clear difference in categorization slopes, whereas Spanish–English bilinguals showed shallower categorization slopes in the wide compared to the narrow condition, but no reliable difference in fixation patterns. Overall, the findings demonstrate that distributional learning extends to bilingual contexts but is constrained by language structure, language experience, and the measure used to index perception. Gradiency is malleable in bilingual listeners, but its adaptation is constrained by the acoustic–phonetic structure of the languages involved. Specifically, gradiency may be more malleable in Spanish–Basque bilinguals than in Spanish–English bilinguals, potentially due to the smaller and more overlapping VOT ranges in Spanish and Basque relative to English. The results also suggest that mouse-click responses and fixation proportions may differentially reflect consistency and gradiency, respectively.

Experiment 2 focuses on the functional roles of gradiency and consistency. Prior research has shown that gradiency facilitates speech perception flexibility (i.e., recovery from misperceptions) in L1, but whether and how gradiency contributes to speech perception flexibility in L2 remains unknown. In addition, although previous research has demonstrated multiple advantages of consistency, its role in initial lexical activation (i.e., early activation of word candidates) and speech perception flexibility has not been tested. To address these gaps,

this experiment investigated the roles of gradiency and consistency in spoken-word recognition among Spanish (L1)–English (L2) bilinguals. Gradiency and consistency were assessed using a Visual Analogue Scale (VAS), in which participants indicated their perceptual judgments of stop consonants (/ba/–/pa/) presented along a continuum manipulated in 7–8 ms steps of VOT, using a continuous response line; gradiency and consistency were quantified via curve fitting of the resulting response functions. Initial lexical activation and speech perception flexibility were assessed using an eye-tracking Visual World Paradigm task, in which listeners were presented with misleading stimuli (e.g., *peachball*, *beachpit*) on some trials. This design allowed us to capture both initial activation of VOT-compatible lexical candidates at word onset and recovery dynamics following disambiguation. Seventy Spanish–English bilinguals completed these tasks in both languages. Participants also completed additional tasks assessing cognitive and linguistic factors potentially related to gradiency, including inhibitory control, working memory, and an English picture-naming task to index English proficiency.

Replicating previous findings in L1 English, gradiency facilitated speech perception flexibility in L1 Spanish. In contrast, gradiency did not facilitate L2 speech perception; instead, a different pattern emerged: participants relied more heavily on lexical (top–down) than subphonemic (bottom–up) information, given less robust category representations in L2. These findings suggest that the functional role of gradiency in L1 versus L2 speech perception is modulated by differential reliance on bottom–up versus top–down information. In contrast, gradiency was not associated with initial lexical activation. Nonetheless, listeners with higher consistency exhibited stronger initial lexical activation for acoustically compatible word candidates across /b/ and /p/ in L1 Spanish and for /p/ in L2 English. However, consistency was not associated with speech perception flexibility. These findings indicate that speech perception consistency primarily supports early lexical activation during

spoken-word recognition, while playing a more limited role in later processes involved in revising initial misinterpretations. The divergent functional roles of gradiency and consistency further suggest that they reflect dissociable mechanisms. Moreover, a positive correlation between L1 and L2 gradiency was observed only among higher-proficiency listeners, whereas consistency showed greater cross-language stability overall. Taken together, these findings suggest that consistency likely reflects the stability of cue-to-category mappings and constitutes a more general trait than gradiency.

Experiment 3 tests whether gradiency and consistency are linked between perception and production. Specifically, it examines whether listeners with higher perceptual gradiency also demonstrate greater imitation accuracy or gradiency in speech production, and whether listeners with higher perceptual consistency exhibit greater consistency in production. To address these questions, 80 Spanish speakers completed a VAS task measuring perceptual gradiency and consistency along a Spanish /ba/–/pa/ continuum of VOT. The same participants also performed an explicit imitation task using the same stimuli. Imitation accuracy was quantified as the inverse of the difference between stimulus VOT and imitated VOT, taking baseline production of /b/ and /p/ into account. Two additional measures were derived: a simple imitation accuracy measure that did not account for baseline production, and an imitation gradiency measure based on the same curve-fitting procedure used in the VAS task. Correlation and regression analyses revealed only weak relationships between perceptual gradiency and all imitation measures (imitation accuracy, simple imitation accuracy, and imitation gradiency), as well as between perceptual consistency and imitation consistency, even after accounting for other factors related to speech perception and imitation, including openness to experience, attention switching, articulatory ability, and musical perception skills. These findings suggest that models proposing strong perception–production coupling may not extend to the subphonemic level.

Together, these findings demonstrate that gradiency and consistency reflect distinct yet complementary properties of speech perception. They highlight the importance of incorporating language-specific structure and listener experience into models of spoken-word recognition. Speech perception is neither uniformly gradient nor uniformly consistent, but is dynamically shaped by language structure and experience.

Resumen en Castellano

Esta tesis investiga la naturaleza multidimensional de la percepción subfonémica del habla, centrándose en dos constructos: la gradiencia (gradiency), definida como la sensibilidad a las diferencias acústicas dentro de un mismo fonema, y la consistencia (consistency), definida como la estabilidad de las respuestas ante un mismo sonido del habla a lo largo de repeticiones. Aunque investigaciones previas han mostrado que los oyentes son, en general, gradientes (gradient listeners), existen diferencias individuales sustanciales. Se observan diferencias individuales similares en la consistencia. Sin embargo, las fuentes de estas diferencias individuales, la maleabilidad de la gradiencia y la consistencia, y las consecuencias funcionales de estos constructos siguen siendo en gran medida desconocidas.

Con este fin, a través de tres experimentos, esta tesis examina la maleabilidad de la gradiencia y los roles funcionales de la gradiencia y la consistencia, con especial énfasis en los contrastes de sonoridad bilabial (/b/–/p/) en español, euskera e inglés, utilizando el tiempo de inicio de la sonoridad (voice onset time, VOT) como principal pista acústica (acoustic cue). Estos experimentos también arrojan luz sobre las fuentes y las características de ambos constructos. Al centrarse en oyentes multilingües e incorporar múltiples lenguas, esta tesis amplía una literatura que se ha basado en gran medida en oyentes monolingües de inglés.

El Experimento 1 se centra en la maleabilidad de la gradiencia mediante un paradigma de aprendizaje distribucional (distributional learning paradigm). Incluye dos subexperimentos que evalúan si los oyentes bilingües ajustan su percepción del contraste /b/–/p/ en función de la varianza de las distribuciones de VOT. Setenta y dos bilingües español–euskera (Experimento 1.1) y 72 bilingües español–inglés (Experimento 1.2) fueron expuestos a distribuciones de VOT estrechas o amplias en cada lengua, operacionalizadas mediante la distribución de la frecuencia de ensayos a lo largo de los distintos pasos de VOT, con varianza igual o mixta entre lenguas. El aprendizaje distribucional se evaluó mediante dos

medidas: la pendiente de las funciones de categorización (categorization functions) basada en respuestas de clic con el ratón y la proporción de fijaciones hacia competidores fonológicos (phonological competitors) en una tarea de tipo Visual World Paradigm. Estudios previos con oyentes monolingües de inglés han mostrado que las distribuciones más amplias producen pendientes de categorización más planas y un mayor número de fijaciones a competidores en comparación con las distribuciones estrechas, lo que indica sensibilidad a la varianza distribucional (distributional variance). Sin embargo, en el presente estudio este patrón no se observó de manera consistente. La exposición a distribuciones de VOT más amplias produjo pendientes más planas y un aumento de las fijaciones a competidores en algunas condiciones, pero la magnitud de estos efectos varió en función del par de lenguas y de la medida utilizada para indexar la percepción del habla (speech perception). En concreto, los bilingües español–euskera mostraron un aumento de fijaciones a competidores en la condición amplia en comparación con la estrecha, pero sin diferencias claras en la pendiente de categorización, mientras que los bilingües español–inglés mostraron pendientes más planas en la condición amplia, pero sin diferencias fiables en los patrones de fijación. En conjunto, los resultados demuestran que el aprendizaje distribucional se extiende a contextos bilingües, aunque está condicionado por la estructura lingüística, la experiencia lingüística y la medida empleada. La gradiencia es maleable en oyentes bilingües, pero su adaptación está restringida por la estructura acústico-fonética (acoustic–phonetic structure) de las lenguas implicadas. En particular, la gradiencia puede ser más maleable en bilingües español–euskera que en bilingües español–inglés, posiblemente debido a los rangos de VOT más reducidos y solapados en español y euskera en comparación con el inglés. Los resultados también sugieren que las respuestas de clic y las proporciones de fijación pueden reflejar de manera diferencial la consistencia y la gradiencia, respectivamente.

El Experimento 2 se centra en los roles funcionales de la gradiencia y la consistencia. Investigaciones previas han mostrado que la gradiencia facilita la flexibilidad en la percepción del habla (speech perception flexibility; es decir, la recuperación tras percepciones erróneas, recovery from misperceptions) en L1, pero se desconoce si y cómo contribuye a esta flexibilidad en L2. Además, aunque se han documentado diversas ventajas de la consistencia, su papel en la activación léxica inicial (initial lexical activation; esto es, la activación temprana de candidatos léxicos, early activation of lexical candidates) y en la flexibilidad perceptiva (speech perception flexibility) no ha sido examinado. Para abordar estas cuestiones, este experimento investigó los roles de la gradiencia y la consistencia en el reconocimiento de palabras habladas (spoken-word recognition) en bilingües español (L1)–inglés (L2). La gradiencia y la consistencia se evaluaron mediante una escala analógica visual (Visual Analogue Scale, VAS), en la que los participantes indicaban sus juicios perceptivos sobre consonantes oclusivas (/ba/–/pa/) presentadas a lo largo de un continuo (continuum) manipulado en pasos de 7–8 ms de VOT, utilizando una línea de respuesta continua; ambas se cuantificaron mediante el ajuste de curvas (curve fitting) de las funciones de respuesta resultantes. La activación léxica inicial (initial lexical activation) y la flexibilidad perceptiva (speech perception flexibility) se evaluaron mediante una tarea de seguimiento ocular del tipo Visual World Paradigm, en la que los participantes eran expuestos a estímulos engañosos (misleading stimuli; por ejemplo, peachball, beachpit) en algunos ensayos. Este diseño permitió captar tanto la activación inicial de candidatos léxicos compatibles con el VOT al inicio de la palabra (word onset) como la dinámica de recuperación tras la desambiguación (recovery dynamics following disambiguation). Setenta bilingües español–inglés completaron estas tareas en ambas lenguas. Además, realizaron tareas adicionales que evaluaban factores cognitivos y lingüísticos potencialmente relacionados con la gradiencia, como el control inhibitorio (inhibitory control), la memoria de trabajo (working memory) y

una tarea de denominación de imágenes en inglés (picture naming task) para medir la competencia en esta lengua (language proficiency).

Replicando hallazgos previos en L1 inglés, la gradiencia facilitó la flexibilidad perceptiva en L1 español. En contraste, la gradiencia no facilitó la percepción en L2; en su lugar, emergió un patrón diferente: los participantes dependieron en mayor medida de información léxica (top–down information) que de información subfonémica (bottom–up information), dado que las representaciones categoriales (category representations) en L2 son menos robustas. Estos resultados sugieren que el papel funcional de la gradiencia en L1 frente a L2 está modulado por la distinta dependencia de información de abajo arriba (bottom–up) frente a arriba abajo (top–down). Por el contrario, la gradiencia no se asoció con la activación léxica inicial (initial lexical activation). No obstante, los participantes con mayor consistencia mostraron una activación léxica inicial más fuerte de candidatos compatibles acústicamente con /b/ y /p/ en L1 español y con /p/ en L2 inglés. Sin embargo, la consistencia no se asoció con la flexibilidad perceptiva (speech perception flexibility). Estos hallazgos indican que la consistencia en la percepción del habla contribuye principalmente a la activación léxica temprana durante el reconocimiento de palabras habladas, desempeñando un papel más limitado en los procesos posteriores de revisión de interpretaciones iniciales (revision of initial misinterpretations). Los distintos roles funcionales de la gradiencia y la consistencia sugieren además que constituyen mecanismos disociables (dissociable mechanisms). Asimismo, se observó una correlación positiva entre la gradiencia en L1 y L2 únicamente en participantes con mayor nivel de competencia, mientras que la consistencia mostró una mayor estabilidad entre lenguas (cross-language stability). En conjunto, estos resultados sugieren que la consistencia refleja probablemente la estabilidad de las correspondencias entre señales acústicas y categorías fonológicas (cue-to-category mappings), constituyendo un rasgo más general que la gradiencia.

El Experimento 3 examina si la gradiencia y la consistencia están relacionadas entre percepción y producción (perception–production link). En concreto, analiza si los oyentes con mayor gradiencia perceptiva también muestran mayor precisión o gradiencia en la imitación del habla (speech imitation), y si los oyentes con mayor consistencia perceptiva presentan mayor consistencia en la producción. Para ello, 80 hablantes de español completaron una tarea VAS que medía la gradiencia y la consistencia perceptivas a lo largo de un continuo /ba/–/pa/ manipulado en términos de VOT. Los mismos participantes realizaron también una tarea explícita de imitación con los mismos estímulos. La precisión de la imitación (imitation accuracy) se cuantificó como el inverso de la diferencia entre el VOT del estímulo y el VOT producido, teniendo en cuenta la producción basal (baseline production) de /b/ y /p/. Además, se derivaron dos medidas adicionales: una medida simple de precisión de imitación (simple imitation accuracy) sin ajuste por producción basal y una medida de gradiencia en la imitación (imitation gradiency) basada en el mismo procedimiento de ajuste de curvas utilizado en la tarea VAS. Los análisis correlacionales y de regresión revelaron relaciones débiles entre la gradiencia perceptiva y todas las medidas de imitación, así como entre la consistencia perceptiva y la consistencia en la producción, incluso tras controlar otros factores relacionados con la percepción y la imitación del habla, como la apertura a la experiencia (openness to experience), la flexibilidad atencional (attention switching), la habilidad articulatoria (articulatory ability) y la percepción musical (musical perception). Estos resultados sugieren que los modelos que proponen un fuerte acoplamiento entre percepción y producción (strong perception–production coupling) pueden no ser aplicables al nivel subfonémico (subphonemic level).

En conjunto, estos hallazgos demuestran que la gradiencia y la consistencia constituyen propiedades distintas pero complementarias de la percepción del habla. Asimismo, ponen de relieve la importancia de incorporar la estructura específica de cada

lengua y la experiencia del oyente en los modelos de reconocimiento de palabras habladas (spoken-word recognition models). La percepción del habla no es ni uniformemente gradual ni uniformemente consistente, sino que está dinámicamente moldeada por la estructura lingüística y la experiencia.

Table of Contents

Acknowledgements v

Abstract vii

Resumen en Castellano xii

Table of Contents xviii

1 Introduction 1

1.1 From Categorical Perception to Gradiency 2

1.2 The Recent Rise of Consistency 11

1.3 The Link between Gradiency and Consistency and Their Characteristics 14

1.4 The Present Dissertation 17

2 Malleability of Gradiency 20

2.1 Introduction 20

2.2 Method of Experiment 1.1 28

2.2.1 Participants 28

2.2.2 Task Overview 29

2.2.3 Stimuli 29

2.2.4 Procedure 32

2.2.5 Eye-tracking Recording and Analysis 34

2.2.6 Statistical Framework 34

2.2.7 Statistical Analyses 36

2.3 Results of Experiment 1.1 39

2.4 Interim Discussion 47

2.5 Method of Experiment 1.2 49

2.5.1 Participants 49

2.5.2 Task Overview 50

2.5.3 Stimuli 50

2.5.4 Procedure 52

2.5.5 Eye-tracking Recording and Analyses 53

2.5.6 Statistical Analyses 53

2.6 Results of Experiment 1.2 54

2.7 Results of the Combined Analysis 58

2.8 Exploratory Analysis 60

2.9 Discussion 61

3 Functions of Consistency and Gradiency: Initial Lexical Activation and Speech Perception Flexibility 69

3.1 Introduction 70

3.2 Method 78

3.2.1 Participants 78

3.2.2 Task Overview 79

3.2.3 VAS 80

3.2.4 VWP 81

3.2.5 English Picture Naming Task 87

3.2.6 Spatial Stroop Task 88
3.2.7 Corsi Block-Tapping Task 88
3.2.8 Online Questionnaire and Training 89
3.2.9 Order of Tasks 90
3.2.10 Extracting Speech Perception Gradiency and Consistency 91
3.2.11 Statistical Analyses 92
3.3 Results 93
3.3.1 Preliminary Analyses: VAS 93
3.3.2 Preliminary Analyses: VWP 99
3.3.3 Is L2 Gradiency/Consistency Related to L1 Gradiency/Consistency? 110
3.3.4 Gradiency Effects on Initial Lexical Activation and Speech Perception Flexibility in L1 and L2 112
3.3.5 Consistency Effects on Initial Lexical Activation and Speech Perception Flexibility in L1 and L2 121
3.4 Discussion 125
3.4.1 Differential Effects of Gradiency Shaping L1 and L2 Speech Perception Flexibility 126
3.4.2 Consistency Facilitates Initial Lexical Activation in Both L1 and L2, but Not Speech Perception Flexibility 130
3.4.3 Cross-Language Relationships in Gradiency and Consistency 132
3.4.4 Conclusion 134
4 Functions of Consistency and Gradiency: Perception–Production Alignment 135
4.1 Introduction 136
4.2 Method 143
4.2.1 Participants 143
4.2.2 Task Overview 144
4.2.3 VAS 145
4.2.4 Imitation Task 146
4.2.5 Temporal Order Judgment Task 146
4.2.6 Diadochokinetic Rate (DDK) Task 147
4.2.7 Corsi Block-Tapping Task 148
4.2.8 Mini-PROMS Test 148
4.2.9 Questionnaire 149
4.2.10 Extracting Speech Perception Gradiency and Consistency 149
4.2.11 Extracting Imitation Measures 149
4.2.12 Extracting Articulatory Ability 151
4.2.13 Statistical Analyses 152
4.3 Results 153
4.4 Discussion 164
5 General Discussion 170
5.1 Summary of Results 170
5.2 Malleability of Gradiency 171
5.3 Functional Roles of Gradiency and Consistency 173
5.4 Sources of Consistency and Gradiency 176
5.5 Characteristics of Gradiency and Consistency 181
5.6 Cross-Linguistic Generality of Gradiency and Consistency 183
5.7 Conclusions 185
List of Figures 188

List of Tables *190*

References *191*

Appendices *227*

Appendix 1: Power Analyses 227

Appendix 1.1: Experiment 1 227

Appendix 1.2: Experiment 2 227

Appendix 1.3: Experiment 3 227

Appendix 2: Tables for Experiment 1 228

Appendix 3: Figures for Experiment 2 234

Appendix 4: Tables for Experiment 2 238

Appendix 5: Figure for Experiment 3 248

Appendix 6: Tables for Experiment 3 249

1 Introduction

Speech perception is a remarkable feat of the human mind. In contrast to reading, which allows the visual input to remain available for reinspection, listening unfolds in real time: the acoustic signal is transient, and listeners cannot "rehear" past information. As a result, speech perception requires rapid and efficient mechanisms to extract linguistic meaning from a continuously changing input. From such a rapidly changing acoustic waveform, listeners are able to extract linguistic structure, recognize words and sentences, and understand each other effortlessly in everyday communication. Yet beneath this apparent ease lies an extraordinary computational challenge: the speech signal is highly variable, influenced by talker differences, speaking rate, and contextual factors (e.g., Allen & Miller, 2004; Liberman et al., 1967; Lisker & Abramson, 1964; Miller et al., 1986).

To overcome this variability, the perceptual system must strike a delicate balance between flexibility and stability (e.g., Baese-Berk, 2018; Caudrelier et al., 2024; Chiu et al., 2025; Symons et al., 2026; Theodore et al., 2015). For example, when an English listener converses with a Spanish speaker in English, the English listener may perceive an English /p/ that sounds more like English /b/ due to the influence of Spanish phonetics. Perceptual processing must be flexible enough to adapt to this variation to understand the speaker, yet stable enough to recognize /p/ and /b/ as distinct categories in the listener's own language. Flexibility allows listeners to adapt to variability in the signal, while stability ensures consistent mapping from variable acoustics to linguistic categories (Samuel & Kraljic, 2009; Sjerps & McQueen, 2010). Listeners differ in the extent to which they exhibit these two properties, with this balance reflected in two key constructs at the subphonemic level: (1) gradiency, defined as sensitivity to within-category variation, and (2) consistency, defined as the stability of listeners' perception of the same stimulus (reflected in the consistency of their responses to repeated presentations of the same stimulus).

While it is now well established that listeners are generally gradient (e.g., Kapnoula et al., 2017; Massaro & Cohen, 1983; McMurray et al., 2002; Toscano et al., 2010; see McMurray, 2022, for a review) and that the sources of gradiency are partially understood (e.g., Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; Kutlu et al., 2024), less is known about whether gradiency can be modified through experience and what functions it may serve in multilingual processing. Even less work has addressed consistency, a construct that has recently received attention, but has not yet been incorporated into current models of speech processing, leaving its sources and functions largely unknown. Importantly, most previous studies have focused on monolingual English speakers, leaving it unclear how these phenomena manifest in other languages and in multilingual contexts. To address these gaps, this dissertation examines the sources, malleability, functions, and characteristics of gradiency and consistency, focusing on the bilabial stop voicing contrast (/b/–/p/) across Spanish, Basque, and English.

1.1 From Categorical Perception to Gradiency

Speech perception rarely depends on a single cue. When processing spoken language, listeners must map a complex set of acoustic cues onto discrete speech categories. For instance, the perception of voicing contrasts (e.g., /b/ vs. /p/) can be influenced by up to 16 co-varying acoustic dimensions (Lisker, 1986). Among these, Voice Onset Time (VOT), the delay between the release of a stop consonant and the onset of vocal fold vibration for the following vowel, serves as the primary acoustic cue that listeners use to distinguish voiced from voiceless stop consonants in many languages. In English, VOTs of roughly 0–10 ms are perceived as /b/, whereas those of 30–70 ms are perceived as /p/, with the category boundary typically occurring between 10 and 25 ms (Allen & Miller, 1999; Lisker & Abramson, 1964). However, because cue values vary systematically with talker identity (Allen & Miller, 2004), speaking rate (Miller et al., 1986; Miller & Volaitis, 1989), and other contextual factors, the

same cue value does not always correspond to the same category, an observation known as the *lack of invariance* problem.

Early views of speech perception suggested that listeners overcome this problem by largely ignoring within-category information. This idea is based on the empirical phenomenon known as *categorical perception*, the finding that two speech sounds from the same phoneme category are often less distinguishable compared to an acoustically equidistant pair of sounds that map onto different phonemes (Liberman et al., 1957, 1961; Repp, 1984; M. E. H. Schouten & van Hessen, 1992). This well-replicated finding lent support to the idea that within-category differences are largely ignored, and speech sounds are perceived categorically. For example, assuming a boundary at 25 ms, an English /b/ sound with a VOT of 0 ms should be essentially indistinguishable from one with a VOT of 20 ms, whereas a /b/ with VOT of 10 ms and a /p/ with VOT of 30 ms should be more easily distinguished, despite the equal acoustic distance between the two pairs.

Despite the early popularity of the categorical perception account, advances in psycholinguistics have challenged this perspective by showing that listeners are sensitive to within-category information, suggesting that speech categorization is fundamentally *gradient* (see McMurray, 2022, for a review). According to the gradient view of speech perception, listeners can detect subtle, continuous variation in phonetic cues rather than perceiving speech as strictly categorical. Converging evidence for gradiency comes from behavioral, eye-tracking, electrophysiological, and neuroimaging studies (e.g., Andruski et al., 1994; Bidelman et al., 2013; Blumstein et al., 2005; Carney et al., 1977; Kapnoula et al., 2017; Kapnoula & McMurray, 2021; Marslen-Wilson & Warren, 1994; Massaro & Cohen, 1983; McMurray et al., 2002, 2008, 2014; Miller, 1994; Myers et al., 2008; Ou & Yu, 2022; Pereira et al., 2018; Pisoni & Lazarus, 1974; Pisoni & Tash, 1974; Samuel, 1977, 1982; Sharma et al., 1993; Toscano et al., 2010; Utman et al., 2000). For example, behaviorally, response

times (RTs) for "same" judgments are faster for identical than for acoustically different tokens within a phoneme category (Pisoni & Tash, 1974). Gradiency is also evident in fixation patterns in eye-tracking studies using the Visual World Paradigm (VWP; Tanenhaus et al., 1995), a reliable task that tracks listeners' eye movements to visual referents as they process spoken words (Farris-Trimble & McMurray, 2013), providing a real-time index of lexical activation. When the target word (e.g., *beach*) is acoustically closer to the /b/–/p/ boundary (i.e., more ambiguous), listeners fixate more often on the competitor picture (e.g., *peach*), even though they ultimately select the correct image (*beach*; McMurray et al., 2002, 2014). In EEG research, listeners show continuous modulation of the N1 event-related potential (ERP) component, a marker of early auditory encoding (Sharma & Dorman, 1999; Sharma et al., 2000), with N1 amplitude decreasing (i.e., becoming less negative) as VOT increases (Toscano et al., 2010; see also Frye et al., 2007; Getz & Toscano, 2021; Pereira et al., 2018; Toscano et al., 2018, for other neural evidence). Together, findings from different methodological paradigms provide robust evidence that speech perception operates on graded rather than strictly categorical representations.

If speech perception is inherently gradient, why has the phenomenon of categorical perception appeared so robust and persisted for so many years? One likely explanation lies in the tasks used to study it. Categorical perception is typically defined by three characteristic patterns: (1) a sharp increase in category responses around the boundary in identification tasks, (2) a corresponding peak in discrimination performance for stimulus pairs that cross that boundary, and (3) discrimination performance that remains close to chance within each category (Repp, 1984; Studdert-Kennedy et al., 1970). The two-alternative forced-choice (2AFC) task is the standard identification method used to study phoneme categorization. In a 2AFC task, participants hear a token drawn from a speech continuum varying in small steps between two sounds (e.g., /b/–/p/) and must make a forced-choice judgment (e.g., "ba" or

"pa"). Results are typically interpreted through the assumption that sharp categorization is ideal, and that a shallower psychometric slope indicates noise in the perceptual system.

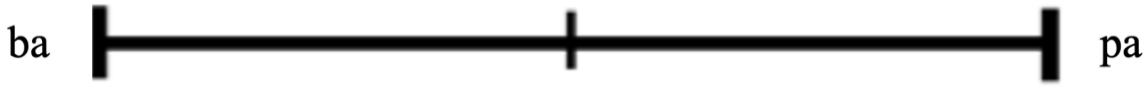
However, this interpretation confounds multiple processes because the 2AFC task does not directly reveal how speech sounds are encoded at early perceptual stages. Consider a listener with a genuinely gradient internal representation; for instance, they perceive a sound 30% along the /b/–/p/ continuum as 30% /p/-like. If this listener consistently chooses the most probable response (e.g., always responding to "ba" until the cue value exceeds 50%), their responses will yield a steep slope, mimicking a categorical pattern despite the underlying gradiency. Conversely, a listener with the same gradient representation might use a probability-matching strategy, responding "pa" about 30% of the time, which would produce a shallower slope. On the other hand, even a listener with truly categorical perception could produce a gradient-looking slope if internal or sensory noise distorted the cue encoding. Thus, in the 2AFC task, it is difficult to determine whether a shallow slope reflects representational gradiency or merely decision or encoding noise. Indeed, previous research has critically examined the limitations of 2AFC tasks (e.g., Apfelbaum et al., 2022; Hary & Massaro, 1982; H. Kim et al., 2025b; McMurray et al., 2008; Pisoni & Tash, 1974).

Similarly, discrimination paradigms are not without limitations. A commonly used approach is the ABX task, where two stimuli (A and B) are presented in sequence, followed by a third (X) that matches either the first or the second. Participants indicate which of the first two matched the third, with chance-level accuracy expected if the first two sounds cannot be discriminated. Classic studies reported strong discrimination across category boundaries but poor performance for within-category pairs (e.g., Liberman et al., 1957; Studdert-Kennedy et al., 1970). However, later research demonstrated that ABX performance reflects not only early auditory encoding but also higher-level categorization processes (e.g., Pisoni & Tash, 1974). The task's memory demands may prompt listeners to rely on category

labels rather than raw perceptual detail (Gerrits & Schouten, 2004; B. Schouten et al., 2003). To mitigate these biases, researchers developed alternative paradigms with lower memory requirements, such as oddball, 4IAX, and 2IFC tasks (Carney et al., 1977; Pisoni & Lazarus, 1974). These methods often reveal considerable within-category sensitivity and no boundary-related discrimination peak (Gerrits & Schouten, 2004; B. Schouten et al., 2003; see Apfelbaum et al., 2022; McMurray, 2022, for related discussion).

To overcome the limitations of the identification and discrimination tasks, researchers have developed a continuous-response task, the Visual Analog Scale (VAS; Figure *1.1*), which captures graded representations more directly. The VAS task was first introduced by Massaro and Cohen (1983). In this task, participants are presented with a continuum labeled, for example, with "ba" at one end and "pa" at the other. Upon hearing a sound from a speech continuum (e.g., a clear "ba," a clear "pa," or something in between), participants indicate its position along the continuum. Unlike the 2AFC task, the VAS task provides a more nuanced measure of perception, capturing the subtle ways in which individuals differentiate speech sounds.

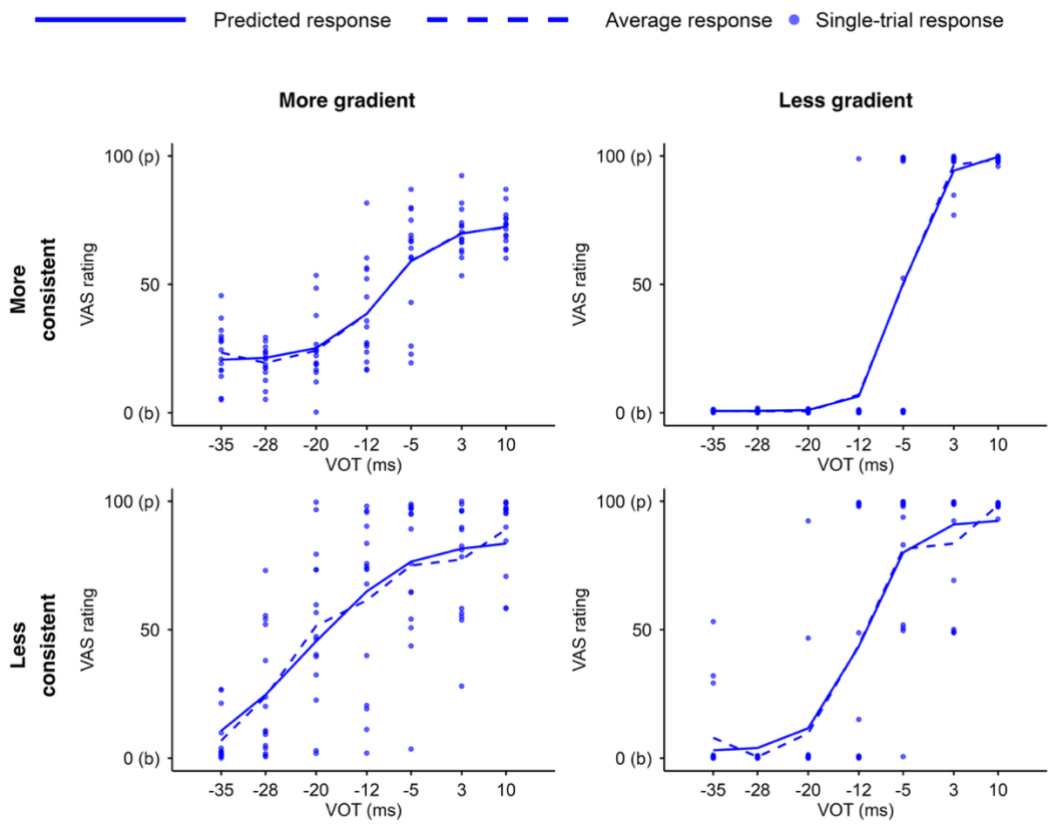
Figure 1.1
Visual Analog Scale (VAS) Task



Although the VAS was developed in the early 1980s, its use has surged in recent years due to the growing interest in individual differences in speech perception. While speech perception is generally gradient, listeners vary in their degree of gradiency (Kapnoula & McMurray, 2021; Kapnoula et al., 2017). These individual differences can be effectively captured with the VAS task, which provides a sensitive yet low-cost and time-efficient alternative to more resource-demanding techniques such as eye-tracking and EEG. Some

individuals are highly attuned to fine-grained acoustic variation, producing more gradient responses characterized by a shallow response function (hereafter, *more gradient listeners*). Others are less sensitive to within-category variation, yielding steeper response functions (hereafter, *less gradient listeners*; Kapnoula et al., 2021; Kapnoula & McMurray, 2021; Kapnoula et al., 2017; Kong & Edwards, 2016). This contrast is illustrated in Figure *1.2*, which presents data from four sample listeners in the Spanish VAS (from Experiment 2 in this dissertation). The left panels show data from two listeners with high gradiency (shallower solid curves), whereas the right panels show data from two listeners with low gradiency (steeper solid curves; consistency will be discussed later in this chapter).

Figure 1.2
Examples of Spanish VAS Data Showing Different Gradiency and Consistency of Four Participants



Note. Each dot represents a participant's response on one trial. The solid line is the fitting curve based on the rotated logistic function, while the dashed line connects the average response at each VOT step for that participant. Note that the x-axis represents the typical range of VOT values for the Spanish /b/–/p/ contrast.

Multiple studies have demonstrated the high reliability and validity of VAS measures. First, gradiency as measured by the VAS task shows strong stability across repeated sessions when the same stimuli are used (Kong & Edwards, 2016; see also Honda et al., 2024a, for relevant discussion), underscoring its test–retest reliability. Importantly, VAS gradiency reflects early stages of speech cue encoding. Evidence for this comes from a study linking VAS gradiency with the N1 ERP component (Kapnoula & McMurray, 2021). As discussed above, this component indexes early auditory encoding. Across listeners, N1 amplitude varied linearly with VOT; however, this relationship was disrupted near the phonemic boundary, but only among listeners who exhibited a more categorical VAS response pattern. Finally, VAS gradiency converges with independent measures derived from other paradigms. Specifically, VAS gradiency positively correlates with lexical gradiency assessed via the VWP and negatively correlates with speech categorization indexed by the P3 ERP component (Kapnoula & McMurray, 2021), which reflects higher-level phonological or lexical categorization processes (Maiste et al., 1995). Collectively, these findings establish the VAS task as a reliable and valid tool for assessing listeners' speech perception gradiency. Moreover, VAS gradiency captures information from both early encoding stages and later categorization processes, linking perceptual and neural representations across multiple levels of processing.

While gradiency is known to originate from early auditory encoding, previous research has examined various factors that may predict individual differences in this construct. However, findings have been somewhat inconsistent. Some studies have reported weak associations between gradiency and cognitive functions such as cognitive flexibility, inhibitory control, and working memory (e.g., Kapnoula et al., 2017; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; Kong & Edwards, 2016). Similarly, H. Kim et al. (2025b) found only weak relationships between VAS gradiency and higher-level cognitive factors

such as autistic traits, anxiety, and impulsivity, suggesting that the connection between gradiency and general cognitive abilities is limited. A more robust predictor appears to be social network diversity: listeners who are exposed to more linguistically diverse environments tend to exhibit greater gradiency (Kutlu et al., 2024). Other contributing factors include domain-general auditory acuity, measured using a temporal order judgment task (Kapnoula & Samuel, 2024), which will also be employed in one of the studies (Experiment 3) in the present dissertation. Listeners with higher auditory acuity tend to show greater speech perception gradiency. Moreover, cumulative language experience, as reflected by age, has also been positively linked to gradiency in adults (Kapnoula & Samuel, 2024), adolescents (McMurray et al., 2018), and school-aged children (Kutlu et al., 2026). Overall, these findings suggest that while gradiency may rely on general auditory mechanisms ("a good ear"), it is also shaped by linguistic experience, which gradually refines the speech perception system and enhances sensitivity to fine-grained acoustic detail.

In sum, while traditional views of speech perception proposed that listeners rely on discrete phonological categories, a wealth of evidence now demonstrates that speech perception is fundamentally gradient. This gradiency, in turn, influences downstream processing at both the lexical (e.g., Andruski et al., 1994; Desmeules-Trudel & Zamuner, 2019, 2023; Marslen-Wilson & Warren, 1994; McMurray et al., 2002) and sentential levels (e.g., Bicknell et al., 2025; Brown-Schmidt & Toscano, 2017; Bushong, 2025; Bushong & Jaeger, 2025; Falandays et al., 2020; Sarrett et al., 2020; cf. Connine et al., 1991). Theoretically, gradiency challenges early views of speech perception as a strictly categorical process. Models such as the Motor Theory (Liberman et al., 1967) proposed that continuous acoustic variation is transformed into discrete phonemic representations, with within-category detail discarded before lexical access. In contrast, modern interactive and connectionist models such as TRACE (McClelland & Elman, 1986) and Shortlist B (Norris

& McQueen, 2008) assume that listeners maintain subphonemic detail throughout processing and use it to guide lexical activation.

Despite these advances, several open questions remain that are central to refining theoretical models of speech perception. Importantly, the recognition of individual differences in gradiency has opened new avenues for research. These include questions related to the sources of gradiency (discussed above), the extent to which gradiency is malleable (i.e., open to change), and the functions it serves in language processing. Examining the malleability of gradiency can determine whether listeners' perceptual systems can be reshaped through exposure and learning, thereby informing debates about plasticity in subphonemic processing. Identifying the functions of gradiency can clarify how retaining fine-grained information supports higher-level processes such as lexical activation. Focusing primarily on English monolinguals, a growing body of research has begun to address these questions (e.g., Bidelman et al., 2025; Fuhrmeister et al., 2023; Gür & McMurray, 2025; Honda et al., 2024a; Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; Kong & Edwards, 2016; Rizzi & Bidelman, 2024; reviewed in subsequent chapters). Importantly, converging evidence suggests that gradiency benefits multiple aspects of speech perception, including integrating multiple acoustic cues (Kapnoula et al., 2017, 2021; D. Kim et al., 2020; Kong & Edwards, 2016; Kong & Kang, 2023; Lee & Park, 2024; Ou et al., 2021; cf. Kong, 2019), improving speech-in-noise performance (Bidelman et al., 2025, 2026; Rizzi & Bidelman, 2024; cf. Kapnoula et al., 2017, 2021), enhancing speech perception flexibility by facilitating the recovery from misperceptions (Kapnoula et al., 2021), and second language (L2) learning (Kapnoula & Samuel, 2024). The present dissertation extends this line of work to a multilingual context, examining a less-studied domain, the malleability of gradiency (Chapter 2), and further exploring its functional roles (Chapters 3 and 4) among Spanish–Basque–English trilinguals.

1.2 The Recent Rise of Consistency

In recent years, a growing body of research has begun to focus on another metric captured by the VAS task, often referred to as *consistency*. Consistency reflects the extent to which listeners provide similar responses across repeated presentations of the same stimulus. Like gradiency, individual differences in consistency have been reported (e.g., Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Honda et al., 2024a; Kapnoula et al., 2017). In the VAS task, multiple instances of each sound are presented and responses to each instance are recorded. For example, in Figure *1.2*, each dot represents a participant's response on an individual trial. Greater dispersion of dots (vertically) around the fitted curve at the same VOT value indicates lower consistency (see lower panel), whereas tighter clustering of dots reflects higher consistency (see upper panel).

Compared to gradiency, much less is known about consistency. For instance, the sources of consistency remain unknown. This uncertainty is reflected in the diverse terminology used across previous studies, such as *categorization consistency* (Carraturo et al., 2025; H. Kim et al., 2025a–f; Rizzi & Bidelman, 2025), *noise* or *variability* (Bidelman et al., 2025; Kapnoula et al., 2017; Rizzi & Bidelman, 2024; Sorensen et al., 2024), *response consistency* (B. Cheng et al., 2025b; Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Honda et al., 2024a, 2024b; Kapnoula & Samuel, 2024), *response variability* (Apfelbaum et al., 2022; B. Cheng et al., 2025a; Kutlu et al., 2022, 2024), and *perceptual consistency* (Fuhrmeister et al., 2023), with some studies even using multiple terms interchangeably. This variation in terminology highlights a lack of consensus regarding whether consistency arises from early perceptual encoding, later categorization processes, decision-level mechanisms, or increased effort or attention. To remain neutral, the present dissertation adopts the general term "consistency."

Since the resurgence of the VAS, a growing body of research has begun to investigate speech perception consistency. Recent studies suggest that it plays an important role not only in speech perception, but also in broader language-related abilities. Specifically, higher consistency has been positively associated with reading and language skills (Carraturo et al., 2025; H. Kim et al., 2025b–d), improved learning of unfamiliar contrasts (Fuhrmeister et al., 2023; Honda et al., 2024a, 2024b), better speech-in-noise perception (Myers et al., 2024b; Rizzi & Bidelman, 2025), more adaptive use of acoustic cues across phonological contrasts (H. Kim et al., 2025a), more accurate novel accent perception (Kutlu, 2025), and faster spoken word recognition of stop consonants (H. Kim et al., 2025f). Collectively, these findings indicate that listeners with higher consistency tend to perform better across multiple aspects of speech perception and language learning.

Potentially, consistency could reflect domain-general factors such as effort, attention, or task engagement. Supporting this possibility, listeners who show higher consistency in the auditory VAS task also tend to exhibit higher consistency in a visual VAS task (H. Kim et al., 2025d), suggesting that some individuals may simply be generally more attentive or reliable participants. However, there is also evidence that speech perception consistency plays a role that cannot be reduced to these factors. For example, H. Kim et al. (2025e) found no correlation between VAS consistency and higher-level factors such as autistic traits, anxiety, or impulsivity, indicating that VAS consistency is not simply driven by broad individual differences in personality or affect. That said, these findings do not rule out the possibility that listeners with higher speech-perception consistency may simply allocate greater attention or effort to the task. More compelling evidence for a speech-specific role comes from findings that auditory consistency, but not visual consistency, predicts language skills, including receptive vocabulary and sentence comprehension, in adults (H. Kim et al., 2025d). Similar relationships between consistency and reading or language ability have also been

observed in children learning English (H. Kim et al., 2025d) and Korean (H. Kim et al., 2025c), with the latter study showing that the association persists even after controlling for nonverbal IQ. Together, these findings suggest that speech-perception consistency may reflect a domain-specific mechanism that contributes uniquely to language learning, beyond general attentional or motivational differences.

Although consistency has been linked to reading and language abilities, the direction of this relationship remains uncertain. As noted by H. Kim et al. (2025b–d), one possibility is that inconsistent categorization acts as a cause, hindering the development of stable phonological categories that support language and reading (i.e., decoding) skills. Alternatively, higher-level structural regularities in language and reading, such as vocabulary knowledge and grapheme–phoneme correspondence, may help stabilize speech perception through a feedback mechanism, termed *auditory/phonological clean-up* (Luthra et al., 2024; Magnuson et al., 2024). Current evidence appears to favor the latter account: VAS responses to words tend to be more consistent than to nonwords, suggesting that lexical knowledge enhances categorization stability (Carraturo et al., 2025). This may reflect a clean-up process in which higher-level linguistic information reduces variability at certain stages of processing, potentially extending to auditory and phonological representations. Nonetheless, the precise mechanisms underlying this relationship remain to be clarified.

Regarding the predictors of consistency, existing evidence is still limited. Age seems to be a relevant one: school-aged children become more consistent with age (Kutlu et al., 2026), but most other factors that have been examined as potential sources of gradiency have not yet been investigated in relation to consistency. Compared to behavioral findings, neural evidence could provide more direct insights into the mechanisms underlying consistency, though current data remain scarce. Fuhrmeister and Myers (2021) reported that greater gyrification in the auditory cortex was associated with less consistent responses, suggesting a

structural basis for consistency. Similarly, steeper 2AFC slopes (which may reflect higher consistency) have been linked to more faithful subcortical speech encoding (Ou & Yu, 2022). Functionally, and most relevantly, Honda et al. (2024b) investigated the relationship between the consistency of the frequency-following response (FFR), an early brainstem response occurring within 5–10 ms after stimulus onset, and VAS consistency. They found no statistically significant association between the two measures. Based on this, the authors proposed that speech perception consistency at the behavioral level may instead reflect later stages of processing, such as categorization or decision-making. However, they also acknowledged that their sample size was small and that a link in consistency between neural and behavioral levels cannot be ruled out. Taken together, the neural sources of consistency remain unresolved.

Although both 2AFC and VAS tasks can measure consistency, most studies have not examined in detail the nature and sources of this theoretical construct. Importantly, variation in language processing can represent signal rather than noise (McMurray et al., 2023). Therefore, understanding the mechanisms underlying consistency is theoretically essential. In addition, consistency has not yet been incorporated into existing models of speech perception, which typically emphasize categorical representations (e.g., Goldinger, 1998; McClelland & Elman, 1986; McMurray & Jongman, 2011). Examining the sources of consistency also carries potential practical implications for language learning and training, given the potential advantages of consistency discussed above. While the sources of consistency are not examined directly, experiments throughout this dissertation shed light on this issue as well as on the characteristics of consistency, which will be discussed further in the General Discussion (Chapter 5).

1.3 The Link between Gradiency and Consistency and Their Characteristics

Figure *1.2* illustrates that individual differences in gradiency and consistency can be largely independent, such that each can vary regardless of the other. Specifically, gradiency reflects the average response pattern across stimuli (i.e., the slope of the mean function), whereas consistency captures trial-by-trial stability in those responses. Accordingly, previous studies have reported mixed findings regarding their relationship: some have found a negative correlation (e.g., Fuhrmeister et al., 2023), others a positive association (e.g., Honda et al., 2024a), and still others no reliable link (e.g., Kapnoula et al., 2017). As noted by Apfelbaum et al. (2022), the two constructs are theoretically orthogonal, as listeners may exhibit high or low consistency irrespective of their level of gradiency. For instance, a shallow slope on the VAS may reflect truly continuous and consistent responses across stimuli with different VOT steps (top-left of Figure *1.2*), but it could also result from inconsistent, noisy responses within each step (bottom-left of Figure *1.2*). In the latter case, the apparent gradiency would not reflect genuine sensitivity to within-category variation, but rather unstable or noisy encoding (Honda et al., 2024a; Kutlu et al., 2022; Sorensen et al., 2024; see Apfelbaum et al., 2022, for discussion). Notably, the 2AFC task cannot distinguish between these two cases; it would simply yield a shallower slope for both. This limitation further highlights the advantage of the VAS in capturing gradiency more directly.

An important distinction between gradiency and consistency is that the former tends to be contrast-specific (e.g., Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; H. Kim et al., 2025d; Myers et al., 2024b; but see Bidelman et al., 2026), whereas the latter appears to be a more general, trait-like characteristic (e.g., Honda et al., 2024a, 2024b; H. Kim et al., 2025d, 2025e). For instance, consistency derived from a 2AFC task was positively correlated with that from the VAS, whereas 2AFC slope was not (Honda et al., 2024a; see also B. Cheng et al., 2025a, for preliminary evidence in lexical-tone processing). Notably, H. Kim et al. (2025e) found that consistency was positively correlated across different continua,

including five vowel pairs, two voicing pairs, and one fricative pair, whereas gradiency showed only weak cross-continuum correlations. A positive correlation was also observed between consistency in a lexical-stress contrast and a stop-voicing contrast (H. Kim et al., 2025a), as well as between word and nonword consistency (Carraturo et al., 2025). Furthermore, H. Kim et al. (2025d) observed a positive correlation between speech-categorization consistency and visual-categorization consistency (measured using a VAS ranging from a red apple to a yellow pear, a paradigm also adopted in one of the studies in this dissertation). In contrast, no such relationship has been reported for gradiency (Kapnoula et al., 2021; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; H. Kim et al., 2025d). Similarly, although VAS gradiency measures for different speech contrasts are often correlated (e.g., Bidelman et al., 2025, 2026; Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Kapnoula & McMurray, 2021; Kong & Kang, 2023), the strength of these correlations appears to depend on the acoustic similarity between continua. For example, stronger correlations are observed between labial and alveolar stop contrasts than between labial stops and fricatives (Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; Myers et al., 2024b). Collectively, these findings suggest that consistency reflects a domain- or contrast-general processing trait, whereas gradiency is more domain- or contrast-dependent.

Importantly, a recent study found a positive correlation between the slope of the 2AFC task and the consistency of the VAS task (Honda et al., 2024a; see also a trend in the same direction in Kapnoula et al., 2017; cf. B. Cheng et al., 2025a), suggesting that the 2AFC slope may reflect consistency rather than categoricity (the inverse of gradiency), contrary to common assumptions. Under the traditional view that categorical perception reflects optimal processing, a steeper slope in the 2AFC task has been considered beneficial. Steeper slopes, interpreted as sharper categorical boundaries, have often been associated with better reading skills, whereas shallower slopes or less categorical listeners have been linked to impaired

reading ability (Godfrey et al., 1981; Joanisse et al., 2000; López-Zamora et al., 2012; Serniclaes et al., 2001, 2005; Werker & Tees, 1987). However, this interpretation has been challenged by recent findings on the functions of gradiency, which suggest that more gradient listeners can in fact exhibit advantages in various domains in language processing (e.g., Bidelman et al., 2025; Kapnoula et al., 2017, 2021; Kapnoula & Samuel, 2024; Kong & Edwards, 2016; Rizzi & Bidelman, 2024). Taken together, these findings imply that the relationship between categorization slope and reading ability may need to be reconsidered. Specifically, differences between better and poorer readers may reflect variation in consistency rather than differences in categoricity, as was long assumed. This reinterpretation aligns with emerging evidence that higher consistency is associated with better reading and language abilities (H. Kim et al., 2025b–d).

Overall, gradiency and consistency may jointly determine how listeners encode, adapt to, and use speech sounds. The present dissertation further examines the relationship between these two constructs to determine how they interact across linguistic contexts. In addition, it explores their broader characteristics: Do gradiency and consistency each correlate between an individual's first language (L1) and second language (L2)?[1] Does consistency reflect a domain-general trait; that is, are more consistent listeners also more stable performers across cognitive tasks beyond speech perception? These questions are addressed throughout the chapters in this dissertation.

1.4 The Present Dissertation

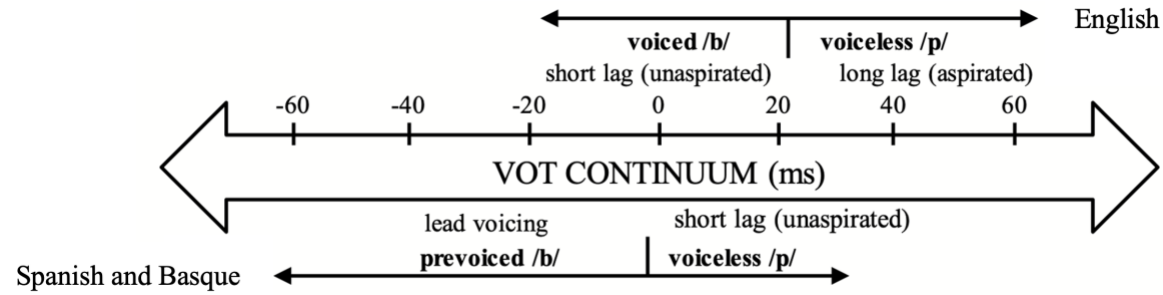
While most studies on gradiency and consistency have focused on L1 speakers of English (and, to a lesser extent, Korean), the present dissertation extends this investigation to L1 Spanish speakers. In some projects, L1/L2 Basque and L2 English are also focused on.

[1] Throughout this dissertation, "L1" and "L2" are used descriptively to refer to relative order of language acquisition. "L2" is used as a general term for later-acquired languages (including second, third, or additional languages).

Examining a broader range of languages is essential for building more generalizable models of language and cognition, particularly in light of the prevailing bias toward English in current research (Blasi et al., 2022). Across studies, participants reside in the Gipuzkoa province of the Basque Country in Spain, where both Spanish and Basque are widely spoken (except for a subset of participants in Experiment 1, who were tested in Connecticut, U.S.A). Most participants have either Spanish or Basque as their L1 or L2 (in varying combinations) and also know English to differing degrees of proficiency. The main focus of this dissertation is on L1 Spanish (Chapters 2–4), with studies including L1/L2 Basque (Chapter 2) and L2 English (Chapters 2 and 3).

This dissertation focuses on the bilabial voicing contrast /b/–/p/, which has traditionally been regarded as one of the most categorically perceived contrasts in classic accounts of speech perception (Liberman et al., 1957). Cross-linguistic comparisons of this contrast are particularly informative because the primary acoustic cue, VOT, differs markedly between English and Spanish. As discussed above, in English, VOTs around 0–10 ms cue /b/, and those near 30–70 ms cue /p/, with the category boundary at approximately 10–25 ms (Allen & Miller, 1999; Lisker & Abramson, 1964). By contrast, Spanish and Basque are true-voicing languages in which the presence of prevoicing distinguishes voiced from voiceless stops. A study conducted in the Gipuzkoa region (i.e., the context of the present dissertation) indicates that typical labial VOTs are approximately /b/ ≈ −80 ms and /p/ ≈ 16 ms for both Peninsular Spanish and Basque (Souganidis et al., 2024). Thus, Spanish and Basque /b/ are prevoiced, and their /p/ acoustically resemble English /b/, with a category boundary around −10 to +10 ms (Benkí, 2005; Williams, 1977b). Figure *1.3* illustrates these cross-linguistic differences in the VOT ranges for bilabial stops. Notably, Spanish /b/ shows greater variability (i.e., a wider VOT distribution) than Spanish /p/ due to prevoicing (Souganidis et al., 2024).

Figure 1.3
VOT Differences between English, Spanish, and Basque /b/–/p/ Contrasts



Note. Adapted from Kapnoula & Samuel (2024); Llama et al. (2010)

Using behavioral and eye-tracking techniques, and adopting an individual-differences approach, the present dissertation reports three experiments that investigate the malleability of gradiency (Chapter 2), and the functional roles of both gradiency and consistency in spoken word recognition (Chapter 3) and in speech production (Chapter 4) across multiple languages. Beyond these main themes, the potential sources of consistency and the characteristics of gradiency and consistency are examined throughout the dissertation, with the key insights brought together in the General Discussion (Chapter 5).

2 Malleability of Gradiency

2.1 Introduction

This chapter presents Experiment 1, which investigates the malleability of speech perception gradiency. The chapter comprises two sub-experiments: Experiment 1.1, focusing on Spanish–Basque bilinguals, and Experiment 1.2, focusing on Spanish–English bilinguals.

While speech perception gradiency has been shown to confer multiple advantages for downstream processing in lexical access, speech-in-noise perception, and language learning (e.g., Bidelman et al., 2025, 2026; Kapnoula et al., 2017, 2021; D. Kim et al., 2020; Kong & Edwards, 2016; Kong & Kang, 2023; Lee & Park, 2024; Ou et al., 2021; Rizzi & Bidelman, 2024), little is known about whether gradiency is trainable or malleable. This question is theoretically important because current models of speech perception do not explicitly account for changes in gradiency through experience or training. It is also practically relevant, given the established benefits of gradiency for perceptual and learning outcomes.

Several early studies on VOT perception investigated the effects of perceptual training on listeners' sensitivity to within-category differences. For example, Carney (1977) employed a same–different discrimination paradigm, in which participants judged whether two sequentially presented sounds were identical or different during training. In addition, Samuel (1977) used an ABX paradigm, where listeners heard two stimuli (A and B) followed by a third stimulus (X) and indicated whether X matched A or B. Results from both studies demonstrated improvements in within-category discrimination performance following training. Other work also provided evidence for plasticity in the speech-perception system using discrimination training with labial stop consonants (e.g., McClaskey et al., 1983; Pisoni et al., 1982). However, it remains unclear whether the discrimination ability measured in these studies reflects the same construct as speech perception gradiency. Moreover, the small sample sizes limited the examination of individual differences in learning outcomes.

While early training studies focused on discrimination accuracy, later work has investigated whether training can modify gradiency measured by VAS. Kapnoula (2016) conducted a training study that directly examined the malleability of speech perception gradiency. Participants completed a pretraining VAS task, a 2AFC training task, and a posttraining VAS task. Critically, participants were assigned to one of three conditions, based on different feedback: bi-dimensional, probabilistic, or unidimensional. The bi-dimensional condition aimed to promote the use of secondary cues, providing feedback based on both F_0 and VOT. The probabilistic condition was designed to encourage a probabilistic mapping between VOT values and phoneme categories. The unidimensional condition reinforced reliance on VOT and discouraged the use of F_0. Results indicated that gradiency increased in all three conditions, with no reliable differences in secondary cue use among the groups. Kapnoula (2016) therefore concluded that gradiency can be enhanced through exposure to a wide range of VOT values, regardless of feedback type.

Preliminary evidence suggests that providing explicit instructions does not reliably enhance gradiency (Giovannone et al., 2024). In this study, participants completed the VAS task twice. The instructions for the second session were manipulated based on each participant's initial gradiency. Participants who were relatively gradient on the first VAS were given instructions encouraging more categorical responses (i.e., focusing on identifying which of the two words they heard), whereas those who were relatively categorical were given instructions encouraging more gradient responses (i.e., focusing on detecting small acoustic differences between sounds). Notably, only listeners who were initially gradient responders appeared to be influenced by the instructional manipulation: gradient listeners could be induced to respond more categorically when instructed to do so, whereas categorical listeners did not appear to become more gradient when instructed to do so.

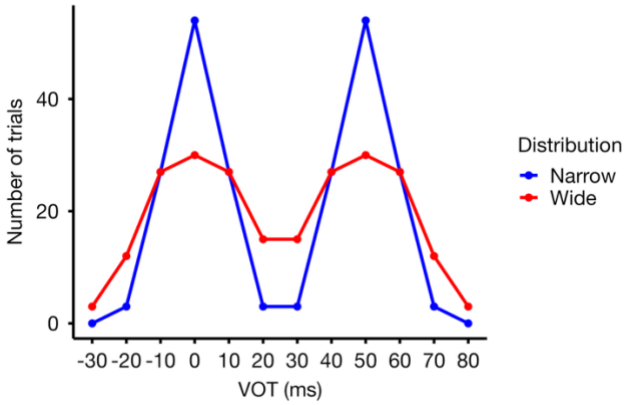
In another study, participants performed a joint 2AFC phoneme identification task with a partner who, unbeknownst to them, was an artificial agent (Nguyen et al., 2024). The responses of the artificial agent were experimentally manipulated with respect to both the category boundary and the slope of the identification function. Convergence was observed for the boundary but not for the slope. The authors attributed the null effect for slope to listeners' prior confidence in the variance of VOT for the two phonemic categories. However, they also noted that the 2AFC task may have been too coarse-grained to capture subtle adjustments in gradiency. Moreover, given the positive correlation between 2AFC slope and VAS consistency reported by Honda et al. (2024a), it is possible that the slope manipulation in this study primarily tapped into consistency, suggesting that consistency, rather than gradiency, may be relatively resistant to change.

Taken together, current evidence suggests that gradiency is unlikely to be modified through training or explicit instruction per se. A complementary line of research has examined whether gradiency can instead emerge implicitly through exposure to natural variability in the speech signal. This idea aligns with the ideal adapter framework (Kleinschmidt & Jaeger, 2015b), which posits that listeners continuously update their internal cue distributions based on experience with variability in the environment. From this perspective, gradiency may reflect an adaptive response to the statistical structure of the input, particularly the amount of variability listeners encounter. This interpretation is consistent with the findings of Kapnoula (2016) mentioned above and with the finding that linguistic diversity facilitates gradiency (Kutlu et al., 2024).

Evidence for implicit changes in gradiency comes from studies using distributional learning paradigms (e.g., Clayards et al., 2008; Nixon et al., 2016), which examine how exposure to variable acoustic input influences cue categorization and perception. In a classic study, Clayards et al. (2008) manipulated the variability of VOTs for the English /b/–/p/

contrast across two groups of participants. Participants were assigned to one of two conditions that differed in VOT distributional variance: a narrow-distribution condition with low variance and a wide-distribution condition with high variance (see Figure *2.1*). Both distributions were bimodal with the same mean VOT, differing only in the degree of variance.

Figure 2.1
Number of Trials at Each VOT Value in the Narrow and Wide Distribution Conditions Used in Clayards et al. (2008)



Note. Dots plotted on the curves mark the presented stimuli and correspond to the x-axis values.

Participants completed a Visual World Paradigm (VWP) task with eye-tracking to assess their sensitivity to the probability distribution of VOT values. On each trial, participants viewed four pictures, heard a word, and selected the image corresponding to the auditory stimulus. The critical stimuli were, for example, *beach* and *peach*, which varied in VOT, while *lace* and *race* served as fillers. When listeners heard *beach*, or a stimulus with an onset closer to /b/, *beach* was the target and *peach* the competitor; the roles were reversed for stimuli with longer VOTs, closer to /p/.

Clayards et al. (2008) found that listeners exposed to the wide-distribution condition showed shallower categorization slopes (i.e., a more gradual increase in responses to /p/-onset pictures across VOT) in mouse-click responses and fixated on the competitor image more frequently than those in the narrow-distribution condition. These results suggest that

increased variability in the acoustic signal heightens perceptual ambiguity and influences how speech is categorized. These findings are consistent with the ideal adapter framework (Kleinschmidt & Jaeger, 2015b), which proposes that listeners are sensitive to the probability distributions of acoustic-phonetic cues and adapt to variation in their statistical properties.

Several studies have replicated the findings of Clayards et al. (2008) using a similar experimental approach (e.g., Gür & McMurray, 2025; Nixon et al., 2016; Theodore & Monto, 2019; Theodore et al., 2020). Nixon et al. (2016) examined the timing of distributional effects in Cantonese and found that these effects emerged rapidly, influencing listeners' initial fixations and persisting throughout the trial. The researchers suggested that the underlying mechanism was not perceptual learning per se but rather a verification process, whereby increased auditory uncertainty prompted participants to seek additional visual information to confirm their decisions. Theodore and Monto (2019) investigated whether speech perception depends on recent or cumulative experience by exposing participants to both narrow and wide distributions. They observed that the order in which distributions were presented did not affect the categorization slope derived from the four-alternative forced-choice (4AFC) task, leading to the conclusion that listeners integrate cumulative exposure to guide phonetic decisions, consistent with the ideal adapter framework (Kleinschmidt & Jaeger, 2015b). Whereas much prior work has focused on the central tendency of VOT and boundary shift (e.g., Chiu et al., 2025; Kleinschmidt & Jaeger, 2015a; Kleinschmidt, 2020; Lindsay et al., 2022; Schreiber et al., 2013; Tan et al., 2022; Tan & Jaeger, 2025), the aforementioned studies demonstrate that listeners are also sensitive to the variance of VOT distributions.

Importantly, a recent study by Gür and McMurray (2025) demonstrated that effects typically attributed to distributional learning may, in fact, reflect increased gradiency in speech perception. Although their study did not employ eye-tracking, it used a distributional

learning paradigm similar to that of Clayards et al. (2008), followed by a VAS task administered after exposure. Crucially, participants exposed to wide VOT distributions exhibited shallower categorization slopes on the VAS, indicating higher gradiency, compared to those exposed to narrow VOT distributions. These preliminary findings suggest that gradiency is not fixed but can be shaped by input variability. Specifically, exposure to greater acoustic-phonetic variability appears to enhance listeners' sensitivity to fine-grained distinctions within phoneme categories.

Although the distributional learning literature has not directly examined the relationship between gradiency and competitor fixations, competitor fixation may serve as an indirect indicator of gradiency. As discussed in Chapter 1, when the target word (e.g., *beach*) is acoustically closer to the /b/–/p/ boundary (i.e., more ambiguous), listeners tend to fixate more frequently on the competitor picture (e.g., *peach*), even though they ultimately select the correct image (*beach*; McMurray et al., 2002, 2014). Importantly, such competitor fixations are associated with gradiency as measured by VAS: listeners with higher gradiency show a more gradual increase in fixations on the competitor as VOT approaches the category boundary (Kapnoula & McMurray, 2021). Thus, proportions of competitor fixations at ambiguous VOTs can provide an informative index of gradiency.

In addition to its relevance for the nature and malleability of speech perception gradiency, the present study is also closely related to the ideal adapter framework. This framework has been supported in a range of monolingual or L1 contexts, including audiovisual speech perception (Bejjanki et al., 2011), selective adaptation and phoneme recalibration (Chládková et al., 2017; Colby et al., 2018; Kleinschmidt & Jaeger, 2012), tone detection in noise and tone duration judgments (Luthra et al., 2025), and the maintenance of subphonemic information over time (Bicknell et al., 2025). However, it has not been

extensively tested in bilingual settings. One notable exception is recent work showing that bilingual listeners can form separate perceptual adjustments for the same acoustic dimension produced by a single speaker across their two languages (Caudrelier et al., 2024). In that study, French–English bilinguals completed two recalibration phases. In the first phase, an ambiguous /s/–/f/ sound replaced either /s/ or /f/ in only one language, depending on group assignment. In the second phase, the same ambiguous sound was embedded in both languages but paired with different intended categories (e.g., /f/ in English and /s/ in French). As a result, listeners simultaneously recalibrated their phoneme boundaries in opposite directions across languages, even though both sets of stimuli were produced by the same talker. This finding suggests that bilinguals can preserve language-specific perceptual mappings even when acoustic information originates from a single speaker. Building on this work, the present study uses a distributional learning paradigm to test whether bilinguals' gradiency reflects the maintenance of distinct representations of VOT variability across their two languages. We conceptualize distributional learning as shaping the probabilistic structure of phonetic representations. Gradiency, in turn, reflects how listeners use this probabilistic structure in speech perception when mapping acoustic cues to categories.

In the present study, we used the same paradigm as in Clayards et al. (2008) to capture listeners' gradiency and to address two research questions. First, we sought to replicate previous findings showing that listeners become more gradient when exposed to speech input with greater VOT variance (e.g., Clayards et al., 2008), extending this effect to a bilingual population. Focusing on Spanish–Basque bilinguals (Experiment 1.1), we included two conditions in which distributional variance of VOT was matched across the two languages (i.e., Spanish-wide/Basque-wide and Spanish-narrow/Basque-narrow; hereafter referred to as the *same-variance groups*). We hypothesized that categorization slopes in mouse-click responses would be shallower and that the proportion of fixations on the

competitor image would be higher in the wide-distribution condition, regardless of whether the language was Spanish or Basque. Based on previous evidence (Kapnoula & McMurray, 2021; McMurray et al., 2002; Gür & McMurray, 2025), shallower categorization slopes and increased competitor fixations (despite ultimately selecting the target) are taken as indicators of more gradient responses.

Our second research question examined whether bilinguals can simultaneously maintain distinct probability distributions of acoustic-phonetic cues for two languages spoken by the same talker, thereby expressing different degrees of gradiency across their two languages. To address this question, we compared two additional conditions in which the distributional variance of VOT differed across the two languages (i.e., Spanish-wide/Basque-narrow and Spanish-narrow/Basque-wide; hereafter referred to as the *mixed-variance groups*). Although the ideal adapter framework (Kleinschmidt & Jaeger, 2015b) was originally developed to account for adaptation to variability across talkers or contexts within a single language, it can be extended to predict that bilingual listeners may maintain separate cue distributions for each language. Under this account, if gradiency reflects sensitivity to distributional information, bilingual listeners should differ in the degree of gradiency expressed across their two languages when the distributional variance differs. Accordingly, we hypothesized that when distributional variance differed across Spanish and Basque, bilinguals would be able to maintain distinct probability distributions for each language, even when both were produced by the same speaker. This would similarly be reflected in a shallower behavioral categorization slope and a higher proportion of fixations on the competitor image in the wide-distribution condition, regardless of whether it corresponded to Spanish or Basque.

Recall that Spanish and Basque employ largely similar VOT ranges (as outlined in Section 1.4 and illustrated in Figure *1.3*), whereas Spanish and English show more distinct

VOT ranges (Figure *1.3*). Because the acoustic distributions of Spanish and Basque overlap substantially, listeners may rely on a largely shared cue distribution rather than maintaining fully separate language-specific distributions. In such cases, language identity provides little additional information for interpreting the acoustic cue, making language-specific distributional adjustments weaker and thereby potentially limiting the extent to which gradiency can adapt to the input when the distributional variance differs across languages. In contrast, the more distinct VOT distributions of Spanish and English may allow listeners to treat language as an informative cue, facilitating language-specific adjustments. To examine whether the degree of VOT-range overlap modulates the malleability of gradiency, we conducted two experiments: one with Spanish–Basque bilinguals (Experiment 1.1) and another with Spanish–English bilinguals (Experiment 1.2). While both experiments targeted the bilabial voicing contrast /b/–/p/ and addressed the same research questions, the degree of VOT overlap differed across the two language pairs. Additional details regarding this VOT-range comparison (our third research question) are presented after Experiment 1.1.

2.2 Method of Experiment 1.1

2.2.1 Participants

We initially planned to recruit 72 participants residing in San Sebastian, Spain, based on a power analysis (Appendix 1.1), resulting in 18 participants per condition (Spanish-narrow/Basque-narrow, Spanish-wide/Basque-wide, Spanish-narrow/Basque-wide, and Spanish-wide/Basque-narrow). Ultimately, 77 participants were recruited because five participants were excluded and replaced because they did not consistently fixate on the target picture when clicking on it (doing so on fewer than 60% of the trials within the 300 ms window preceding their click).

The final sample consisted of 72 Spanish–Basque bilinguals (52 females; 20 males) aged between 18 and 35 years (M = 25.8; SD = 4.7). They demonstrated high proficiency in

both Spanish and Basque, scoring at least 85% on picture-naming tasks in both languages (de Bruin et al., 2017). Participants reported no history of hearing, reading, motor, or speech disorders, and no neurological conditions or head injuries. All had normal or corrected-to-normal vision and no color blindness. All participants spoke Peninsular Spanish. Forty-nine of them spoke the Gipuzkoan dialect of Basque, while 23 spoke Basque without reporting a particular dialect (See Table A2.1 for participant characteristics). These data were collected during participant registration in the Basque Center on Cognition, Brain and Language (BCBL) participant database. The study was approved by the BCBL Ethics Review Board (ref. 240624BC) and adhered to the principles of the Declaration of Helsinki. All participants provided written informed consent and received monetary compensation for their participation.

2.2.2 Task Overview

The task was a VWP eye-tracking experiment. There were four experimental conditions, defined by a 2 (VOT distributional variance: narrow vs. wide) × 2 (language: Spanish vs. Basque) factorial design: Spanish-narrow/Basque-narrow (SNBN), Spanish-wide/Basque-wide (SWBW), Spanish-narrow/Basque-wide (SNBW), and Spanish-wide/Basque-narrow (SWBN). Each participant was assigned to one of these four conditions. The same-variance groups (i.e., SNBN and SWBW) addressed Research Question 1, whereas the mixed-variance groups (i.e., SNBW and SWBN) addressed Research Question 2.

2.2.3 Stimuli

The methods were largely adapted from Clayards et al. (2008). Specifically, for the VWP, three sets of words were used for Spanish and Basque, respectively. Each set contained four words, including one minimal pair beginning with /b/–/p/ (critical stimuli) and one minimal pair serving as fillers. Due to stimulus constraints, the filler contrast was /l/–/r/ in

Spanish and /l/–/x/ in Basque. For example, *bata* [robe], *pata* [paw], *lana* [wool], and *rana* [frog] formed one set in Spanish, while *batu* [to collect], *patu* [fate], *larri* [serious], and *jarri* [to put] formed a corresponding set in Basque. All stimuli within each set had the same number of syllables and identical stress patterns. The number of syllables and stress patterns were also matched across languages. Table *2.1* lists all stimuli used in Experiment 1.1.

Table 2.1
Stimulus Sets Used in Experiment 1.1

Set	Language	Critical minimal pair	English translation	Filler	English translation
1	Spanish	bata – pata	robe – paw	lana – rana	wool – frog
2	Spanish	bajita – pajita	short – straw	lectora – rectora	reader – rector
3	Spanish	válido – pálido	valid – pale	lápida – rápida	tombstone – fast
4	Basque	batu – patu	to collect – fate	larri – jarri	serious – to put
5	Basque	bilatu – pilatu	to search – to pile	lotzea – jotzea	tying – hitting
6	Basque	basarte – pasarte	woodland – passage	lantoki – jantoki	workplace – dining room

All words were recorded by a female Spanish–Basque bilingual speaker from San Sebastian, who spoke Peninsular Spanish and Gipuzkoan Basque. Each word was produced within a carrier phrase—*Es mi* ___ for Spanish or *Nire* ___ for Basque ("It's my ___")—to maintain consistent prosody. To obtain a sufficiently long VOT (up to 45 ms), the speaker was asked to produce the /p/ as in English. Four repetitions of each sentence were recorded. The full recording was denoised in Audacity (Version 2.3.3; Audacity Team, 2019). Target words were then excised from the carrier sentence, and the most natural, clearly articulated token of each word was selected. All tokens were manually inspected for clicks or other unwanted artifacts, which were removed or replaced as needed. Each token was normalized to 70 dB.

The VOT of the critical stimuli (/b/ and /p/) was manipulated to create 12 10-ms steps ranging from −65 to +45 ms using the VOT-synthesis script by Winn (2020; version 33) implemented in Praat (version 6.3.03; Boersma & Weenink, 2023), which affords precise

control over both voicing and aspiration. The script was slightly modified to use the speaker's natural prevoicing extracted from the recording rather than the default prevoicing generated by the script. The rime was excised from a single naturally produced /b/-initial word token to ensure identical segmental context across all items. For negative VOTs, prevoicing segments extracted from that same /b/-initial token were spliced, aligned to pitch periods and zero crossings, into the closure of a naturally produced /p/-initial token. For positive VOTs, the natural voiceless closure was retained, and aspiration noise of the specified duration was appended after the burst using the "Concatenate with overlap" function in Praat. For all stimuli, the fundamental frequency (F_0) was fixed at 170 Hz.[2] Aspiration amplitude remained constant across VOTs. We set the vowel-to-VOT adjustment ratio to 0.65 (default), meaning that only 65% of any increase in VOT duration was compensated by trimming the vowel portion. This procedure extends the progressive cutback-and-replacement method (Andruski et al., 1994) to handle both voiced and voiceless endpoints while preserving smooth, natural transitions. A 50-ms silent interval was added at both onset and offset. Filler items were not manipulated.

Following Clayards et al. (2008), the VOT range was chosen so that participants' average category boundary would fall near the midpoint of the continuum (approximately −10 ms for both Spanish and Basque). The same VOT range and values were used for both languages. For each language, we matched the number of trials per VOT value to those used in Clayards et al. (2008). The number of trials at each VOT value for the narrow- and wide-variance conditions in each language is shown in Table *2.2*.

[2] Previous studies have shown that more gradient listeners rely more heavily on F_0 as a secondary cue for voicing distinctions in L1 English speakers (e.g., Kapnoula et al., 2017). However, Kapnoula and Samuel (2024) reported that Spanish-speaking participants did not appear to use F_0 as a cue in voicing judgments. Similarly, Llanos et al. (2013) found that the influence of F_0 on categorization was minimal, particularly in the prevoicing range. In addition, F_0 has been shown to covary with VOT only across phonological voicing categories, but not within categories (Dmitrieva et al., 2015). Therefore, in the present study, F_0 was held constant and not manipulated across stimuli.

Table 2.2
Number of Trials per VOT Value in the Narrow and Wide Variance Conditions for Spanish and Basque

VOT (ms)	−65	−55	−45	−35	−25	−15	−5	5	15	25	35	45
Narrow	0	3	27	54	27	3	3	27	54	27	3	0
Wide	3	12	27	30	27	15	15	27	30	27	12	3

Note. Trial numbers are identical to those used in Clayards et al. (2008).

Pictures corresponding to each word were taken from the MultiPic database (Duñabeitia et al., 2018) or from online clip art sources. Images were edited to remove irrelevant details and to improve clarity so that they clearly depicted the intended referent. The final images were reviewed and approved by a lab member with extensive experience in VWP studies.

2.2.4 Procedure

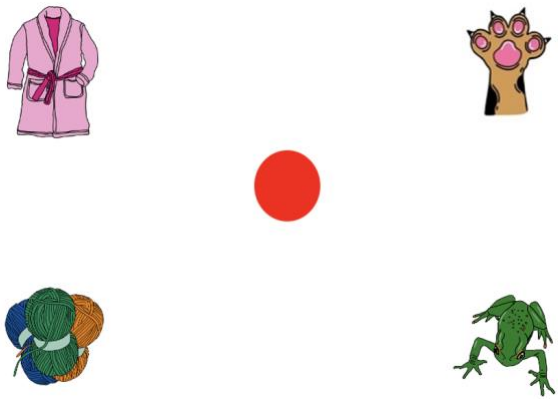
Testing was conducted in a sound-attenuated booth. The experiment was programmed and run using Experiment Builder (version 2.4; SR Research Ltd). Visual stimuli were presented on a 24-inch monitor operating at a resolution of 1024 × 768 pixels. Each display featured four pictures positioned in the four corners of the screen. Each picture measured 240 × 240 pixels, and the distance between the centers of adjacent pictures was 704 pixels horizontally and 448 pixels vertically. The positions of the four pictures were randomized across trials.

Before the main experiment, participants completed a familiarization phase in which they viewed each of the 24 pictures accompanied by their corresponding written labels (12 in Spanish and 12 in Basque), presented one by one in a random order and without audio. This was followed by an eye-tracker calibration procedure and task instructions.

At the beginning of each trial of the main experiment, a red circle appeared at the center of the screen together with four pictures near the corners, two corresponding to the critical stimuli and two to the filler items, representing the words within a given set (see

). After 500 ms, the circle turned blue, and participants clicked on it to hear a word. This brief interval between the appearance of the red circle and the mouse click provided participants with a momentary visual preview of the pictures before the onset of the auditory stimulus, minimizing potential eye movements driven by visual search (Apfelbaum et al., 2021). After hearing the word, participants clicked on the corresponding target picture. Each trial ended once participants clicked on a picture, and the next trial began immediately. Trials from both languages were intermixed and presented in random order. Participants could take a break after every 24 trials and then continue after performing a drift correction, in which they fixated on the central dot and pressed the space bar. The experimenter recalibrated the eye-tracker if the drift correction failed. The main experiment included 456 critical trials (3 critical minimal pairs × 2 languages × 76 repetitions). The three critical minimal pairs in each language were equally distributed across VOT values (all nonzero values in Table *2.2* are divisible by 3). An equivalent number of filler trials was included to obscure the experimental purpose, resulting in a total of 912 trials.

Figure 2.2
VWP Used in Experiments 1.1 and 1.2



Note. This figure shows one of the Spanish stimulus sets: *bata* [robe], *pata* [paw], *lana* [wool], and *rana* [frog].

After the experiment, participants filled out a questionnaire focusing on their listening proficiency in and exposure to Spanish, Basque, and English, their language-switching habits

between Spanish and Basque, and the duration of their musical training (if any; see Table A2.1 for relevant data). The entire session lasted approximately one hour.

2.2.5 Eye-tracking Recording and Analysis

Eye movements were recorded at a sampling rate of 1,000 Hz utilizing the SR Research EyeLink 1000 Plus system, configured in a head-stabilized mode with a chin-rest and a 35-mm lens inside a sound-attenuated booth. The viewing distance between the participant's eyes and the screen was maintained at 70 cm. Overhead lighting and audio volume were kept constant across participants. Auditory stimuli were presented through BeyerDynamic DT–770 Pro 250 Ohm headphones powered by an M-Audio AIR 192|4 amplifier/receiver. The eye tracker was calibrated and validated using the standard nine-point procedure.

The eye-tracking data, recorded from trial onset (the appearance of the blue circle) to the participant's response (mouse click), were automatically parsed into saccades and fixations using the default EyeLink psychophysical parameters. Data preprocessing was conducted using EyelinkAnalysis (Version 5.41; McMurray, 2025). Adjacent saccades and fixations were merged into a single look, beginning at the saccade onset and ending at the fixation offset. For analysis, the data were downsampled to 250 Hz, and each trial was analyzed within a fixed 2,000 ms window relative to stimulus onset. Trials shorter than this duration were extended by holding the final eye position constant, whereas longer trials were truncated, as in previous studies (e.g., Allopenna et al., 1998; McMurray et al., 2002). Picture boundaries were expanded by 100 pixels to accommodate potential noise or head drift in the eye-tracking record. This adjustment did not produce any overlap, maintaining an 8-pixel vertical and 264-pixel horizontal separation between images.

2.2.6 Statistical Framework

Throughout this dissertation, all analyses were conducted in R (version 4.3.2; R Core Team, 2023) via RStudio (version 2023.12.0+369; RStudio Team, 2023). Most statistical analyses were performed within a Bayesian framework, with model specifications described in the relevant dissertation chapters. Here, we summarize the analytical approach that applies across chapters.

Correlation analyses were performed using Spearman's rank-order correlation (ρ) because most variables were not normally distributed and were bounded. Bayes Factors for correlations were obtained using rank-based Bayesian tests implemented in the BayesFactor package (Morey & Rouder, 2024). Bayesian t-tests were also conducted using the same package.

Bayesian linear mixed-effects models were fitted using the brms package (Bürkner, 2021) with a Gaussian distribution. Convergence and sampling quality were assessed using the R-hat statistic (all values ≈ 1), as well as bulk and tail effective sample sizes (ESS > 1,000), indicating good mixing and reliable posterior estimates. All visualizations were produced using the ggplot2 package (Wickham, 2016). Data wrangling and preprocessing were conducted using functions from the tidyverse collection of packages (Wickham et al., 2019).

Bayes Factors (BFs) were computed to evaluate the strength of evidence for individual terms or interactions in each model. A BF = 10 indicates that the evidence for an effect is ten times stronger than for no effect, whereas a BF = 0.1 indicates evidence ten times stronger for the absence of an effect. We interpreted BFs following Kass and Raftery (1995): values between 1–3 indicate weak evidence, 3–10 moderate, 10–30 strong, 30–100 very strong, and > 100 decisive evidence in favor of an effect; conversely, values between 0.33–1 indicate weak evidence, 0.1–0.33 moderate, 0.03–0.1 strong, 0.01–0.03 very strong, and < 0.01 decisive evidence against an effect. Values close to 1 indicate inconclusive evidence.

Throughout the dissertation, effects were considered meaningful when the 95% CrI (credible interval) did not overlap with 0 and the BF indicated at least moderate evidence for the effect.

2.2.7 Statistical Analyses

Before conducting data analyses, all filler trials (32,832 out of 65,664; 50% of the total trials) were excluded from analyses. In addition, 588 experimental trials (1.8% of experimental trials) with RTs exceeding ±3 standard deviations from each participant's mean RT were classified as outliers and removed. Then, 106 trials (0.3% of experimental trials) containing responses to filler items were also excluded. Finally, 143 trials (0.4% of experimental trials) were removed because two participants never selected the picture of *patu* across any VOT for *batu–patu* (the remaining data from these two participants were retained). This resulted in a total of 31,995 trials remaining.

Our primary goal was to examine how distributional variance influenced (a) the slope of the categorization function, based on mouse-click responses, and (b) the proportion of fixations on competitor pictures. Distributional variance (narrow vs. wide) served as the main independent variable. Language (Spanish vs. Basque) and its interaction with distributional variance were additionally included to assess whether the effects of distributional variance generalized across languages. The categorization slope and the proportion of competitor fixations were analyzed as dependent variables in separate models. For Research Question 1 (i.e., whether the malleability of gradiency can be observed in a bilingual context), we analyzed data from the same-variance groups (Spanish-narrow/Basque-narrow and Spanish-wide/Basque-wide), implementing a within-language, between-subjects comparison. For Research Question 2 (i.e., whether gradiency is malleable within individuals across languages), we analyzed data from the mixed-variance groups (Spanish-wide/Basque-narrow and Spanish-narrow/Basque-wide), implementing a within-subject, between-languages comparison.

The first dependent variable, categorization slope, was derived from participants' mouse-click responses in the VWP task. For both languages, the proportion of clicks on the /p/-onset pictures was expected to rise with increasing VOT. Participants whose response patterns did not follow this expected direction would have been excluded; however, all participants showed the expected pattern and were retained for analysis based on this criterion.

The categorization slope parameter was obtained by fitting a four-parameter logistic function to each combination of participant and item based on mean mouse-click responses in MATLAB (version R2022a; The MathWorks Inc., Natick, MA, USA) using the freely available software provided by McMurray (2017, version 48). The model was fit to the average proportion of /p/-onset picture selections (y) across VOT values (x) for each item, minimizing least-squares error under specific constraints (e.g., the crossover constrained within the continuum range and asymptotes bounded). This procedure yielded estimates for the minimum and maximum asymptotes, slope (our primary parameter of interest), and crossover of each participant's response function (see Figure *3.3* for an illustration of these variables in a VAS context). In addition, the fitted category boundary was extracted to determine which picture served as the target and which as the competitor for each item and participant (see details below). Model fits were excellent (average R^2 = .993).

Separate models were fitted to analyze (a) the categorization slope of participants' mouse-click responses and (b) the proportion of fixations on competitor pictures. For the analysis of the categorization slope, the dependent variable was the z-scored slope (within each research-question subset), modeled with a Gaussian likelihood. Fixed effects included distributional variance (narrow vs. wide), language (Spanish vs. Basque), and their interaction. Random-effects structures differed by research question and were selected from a small set of pre-specified candidate structures defined a priori based on the experimental

design, to ensure stable convergence and identifiability (see Appendix 2 for full model formulas). Bayes factors for the main effects of distributional variance and language, as well as their interaction, were computed via bridge sampling by comparing the full model against nested models excluding the relevant term, using the bridgesampling package (Gronau et al., 2020; warp3 method). Models were fitted using weakly informative normal priors on fixed effects and intercepts, and exponential priors on random-effect standard deviations and residual variance.

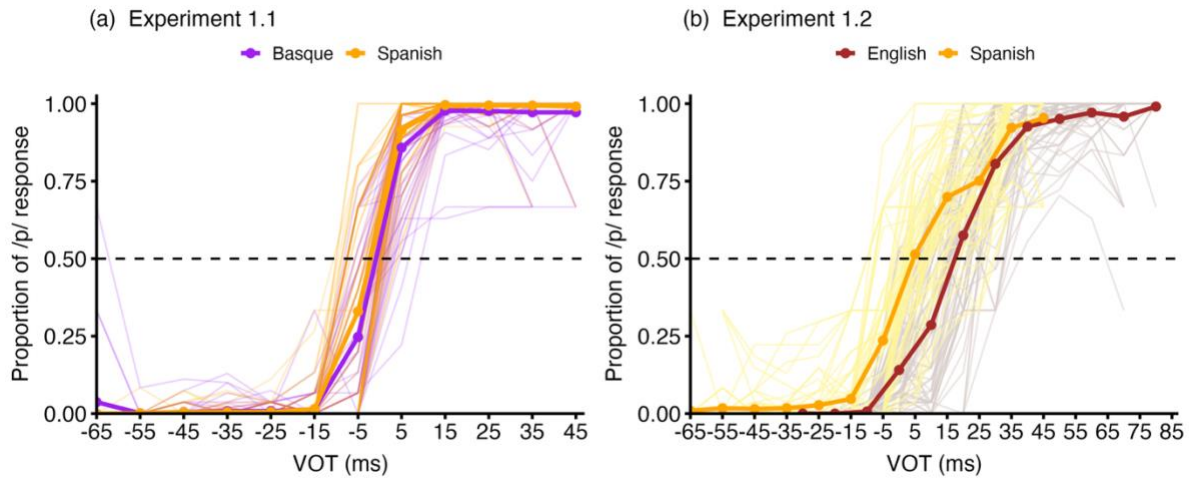
The second dependent variable was the proportion of fixations on competitor pictures (e.g., *pata*) when participants ultimately clicked on the target picture (e.g., *bata*). This proportion was calculated from 250 ms (accounting for a 50 ms onset silence and a 200 ms oculomotor delay; Hallett, 1986; Salverda et al., 2014) to 2,000 ms after stimulus onset. Bayesian mixed-effects models were fitted to analyze the proportion of fixations on the competitor picture in the ambiguous region of the continuum. This analysis focused only on trials in which the VOT was ambiguous for the participant. To capture this ambiguity, we created a variable called relative VOT (*rVOT*; Kapnoula & McMurray, 2021; McMurray et al., 2018). This variable quantified the distance between each participant's category boundary and each stimulus VOT, allowing us to test whether listeners in the wide condition looked more toward the competitor when the stimulus VOT was closer to their boundary. The rVOT was calculated using the category boundary estimated from the logistic function fitted to each participant's mouse-click responses. For each trial, the distance from the boundary was computed by subtracting the item-specific boundary value (estimated for each participant) from the stimulus VOT. Negative rVOT values corresponded to /b/ stimuli, and positive values corresponded to /p/. A total of 564 trials (1.7% of experimental trials) in which participants selected the competitor (i.e., clicking on /p/-onset pictures for negative rVOTs or on /b/-onset pictures for positive rVOTs) were excluded, leaving 31,431 trials for analysis.

For the final analyses, we focused on trials with rVOT values between −15 and +15 ms, as these correspond to the most ambiguous region of the continuum (27.3% of the total observations across both research questions). Because the dependent variable was a proportion bounded between 0 and 1, beta regression with a logit link was used. To accommodate beta regression, competitor-fixation values of 0 and 1 were adjusted to 0.001 and 0.999, respectively. Weakly informative priors included normal(0, 0.5) priors for fixed effects, student-t(3, 0, 0.5) priors for intercepts and random-effect standard deviations, and a gamma(2, 0.1) prior for the precision parameter (φ). Fixed effects, random-effects structures, and Bayes-factor computation followed the same specification as in the slope analysis described above. Full statistical results for Experiment 1 are reported in Appendix 2.

2.3 Results of Experiment 1.1

First, we verified that participants' responses (including all rVOT values and before collapsing by condition) were consistent with expectations. Figure *2.3*a shows individual and average mouse-click responses across VOT values. As shown in the figure, all participants exhibited an increasing proportion of /p/ responses as VOT increased, with category boundaries clustering near 0 ms. Although no explicit time limit was imposed, participants typically responded quickly (Spanish: M = 1,248 ms, SD = 325 ms; Basque: M = 1,423 ms, SD = 409 ms).

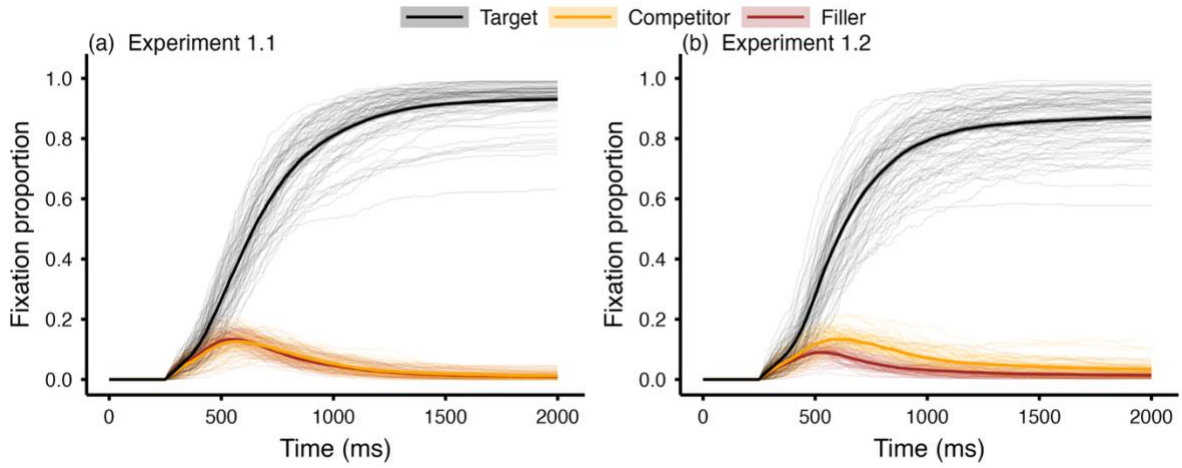
Figure 2.3
Mouse-click Responses of (a) Experiment 1.1 and (b) Experiment 1.2



Note. The y-axis represents the proportion of /p/ responses (i.e., clicks on the /p/-onset picture). Darker curves show the average response across participants, and the lighter curves show individual participants' responses. The black dashed horizontal line indicates the 50% proportion of /p/ responses. Dots plotted on the curves mark the presented stimuli. (a) For both Spanish and Basque, stimuli spanned −55 to +35 ms in the narrow condition and −65 to +45 ms in the wide condition. (b) Spanish stimuli spanned −55 to +35 ms in the narrow condition and −65 to +45 ms in the wide condition. English stimuli spanned −20 to +70 ms in the narrow condition and −30 to +80 ms in the wide condition.

Next, we examined the general fixation patterns on the target, competitor, and filler pictures. Recall that target and competitor pictures were defined according to each participant's item-specific category boundary. Figure *2.4*a presents the individual and average fixation trajectories over time. As shown, participants exhibited increasing fixations on the target as time progressed, whereas fixations on the competitor and each filler picture peaked briefly during the early portion of the trial. Interestingly, the average proportion of fixations on the filler pictures was comparable to that for the competitor, suggesting relatively high overall uncertainty during early processing.

Figure 2.4
Fixation Patterns of (a) Experiment 1.1 and (b) Experiment 1.2

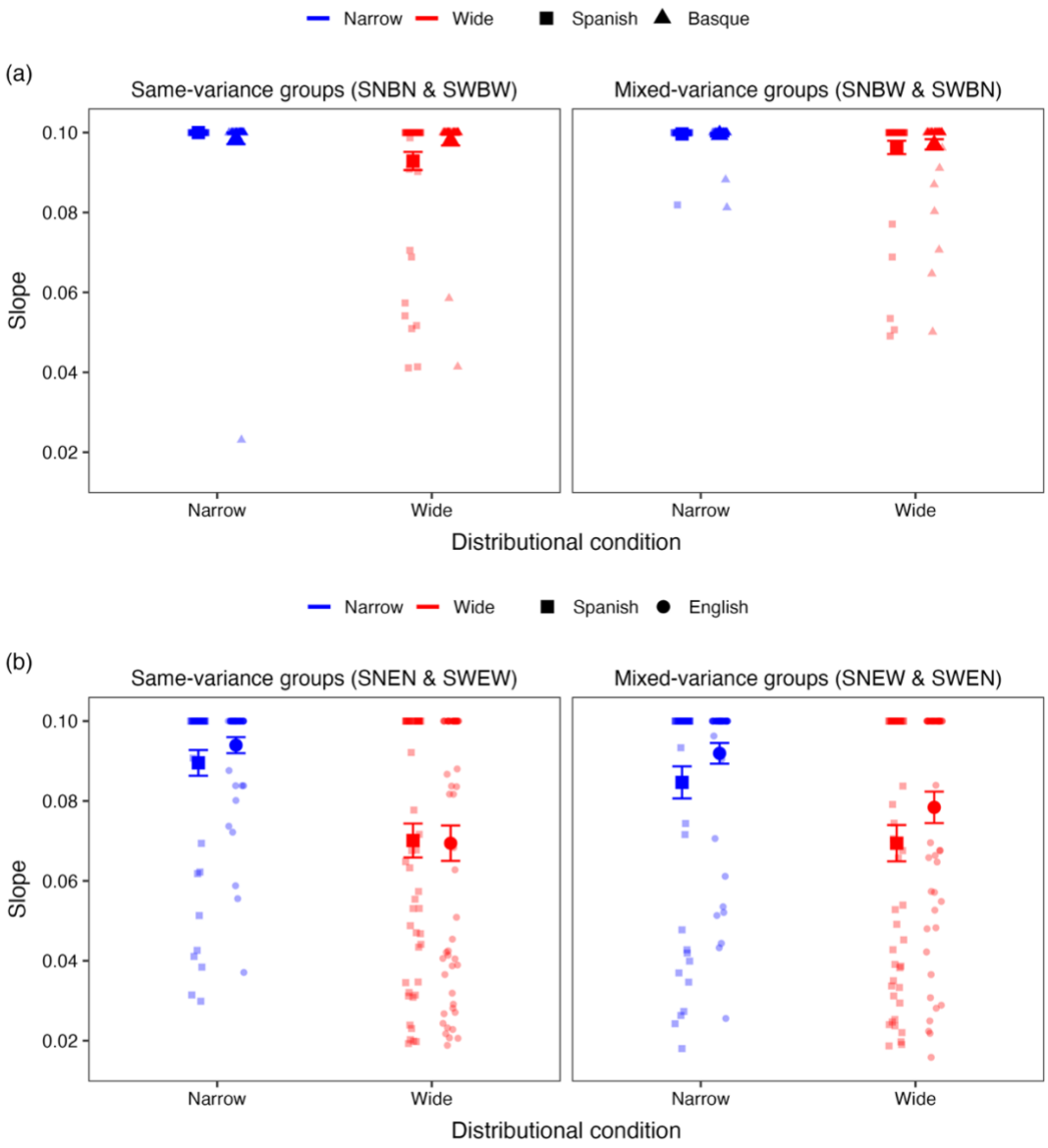


Note. Darker lines show the mean fixations across participants, whereas lighter lines show individual participants' fixations. Fixation proportions during the first 250 ms after stimulus onset were set to zero, corresponding to the 50 ms onset silence and the typical 200 ms oculomotor delay, as this period does not reflect lexical processing.

We now turn to the results addressing our first research question: whether the distributional learning effect on gradiency (i.e., wider distribution driving higher gradiency) can be generalized to a bilingual context. For this question, we focus on the same-variance groups (i.e., SNBN and SWBW). The left panel of Figure *2.5*a shows the participant-by-item categorization slope data based on mouse-click responses for these same-variance groups, comparing the narrow and wide distributional variance conditions in Spanish and Basque (Basque–narrow: $M = 0.099$, $SD = 0.01$; Basque–wide: $M = 0.098$, $SD = 0.01$; Spanish–narrow: $M = 0.100$, $SD = 0.00$; Spanish–wide: $M = 0.093$, $SD = 0.02$). Bayesian mixed-effects modeling revealed moderate evidence in favor of the distributional learning effect: in line with our prediction, categorization slopes were shallower in the wide than in the narrow condition, $\beta = -0.45$, 95% *CrI* (credible interval) [−0.77, −0.13], *BF* (Bayes Factor) = 3.92. This indicates that gradiency is malleable in a bilingual context but modest in magnitude. Consistent with the expectation that slopes would be comparable across languages, there was no reliable main effect of language, $\beta = -0.00$, 95% *CrI* [−0.43, 0.43], $BF = 0.66$, providing weak evidence for the null and indicating similar slopes for Spanish and Basque. Evidence

for an interaction between distributional variance and language was inconclusive, $\beta = 0.34$, 95% *CrI* [−0.76, 0.06], *BF* = 2.15, suggesting no clear difference in distributional learning effects across languages. Full model results are reported in Table A2.2.

Figure 2.5
Participant-by-Item Categorization Slope as a Function of Distributional Condition and Language in (a) Experiment 1.1 and (b) Experiment 1.2



Note. Each small point represents the categorization slope estimate for a given participant and item. Large symbols indicate condition means, with error bars representing ±1 standard error of the mean. A slope of 0.10 is the maximum possible value with the curve fitting algorithm, as justified by the fact that the biggest possible change (i.e., from 0 to 1) between two adjacent steps (10 ms apart) corresponds to a maximum slope of 1/10 = 0.10.

Our second research question is whether gradiency is malleable within individuals across languages. For this question, we focus on data from the mixed-variance (i.e., SNBW and SWBN) groups. The right panel of Figure *2.5*a shows the participant-by-item categorization slope data for these mixed-variance groups, comparing the narrow and wide distributional variance conditions in Spanish and Basque (Basque–narrow: $M = 0.099$, $SD = 0.003$; Basque–wide: $M = 0.097$, $SD = 0.009$; Spanish–narrow: $M = 0.100$, $SD = 0.002$; Spanish–wide: $M = 0.096$, $SD = 0.012$). As the figure shows, categorization slopes were broadly similar across conditions. Consistent with this observation, Bayesian mixed-effects modeling revealed inconclusive evidence for a distributional-variance effect, with categorization slopes only slightly shallower in the wide than in the narrow condition, $\beta = -0.28$, 95% *CrI* [−0.69, −0.17], $BF = 1.41$. Thus, the distributional learning effect was not robust. Consistent with the expectation that slopes would be comparable across languages, there was no reliable main effect of language, $\beta = -0.01$, 95% *CrI* [−0.37, 0.24], $BF = 0.48$, providing moderate evidence for the null and indicating similar slopes for Spanish and Basque. Evidence for the interaction between distributional variance and language was weak, $\beta = -0.31$, 95% *CrI* [−0.57, 0.54], $BF = 0.64$, suggesting that the distributional learning effect did not differ across the two languages. Full model results are reported in Table A2.3.

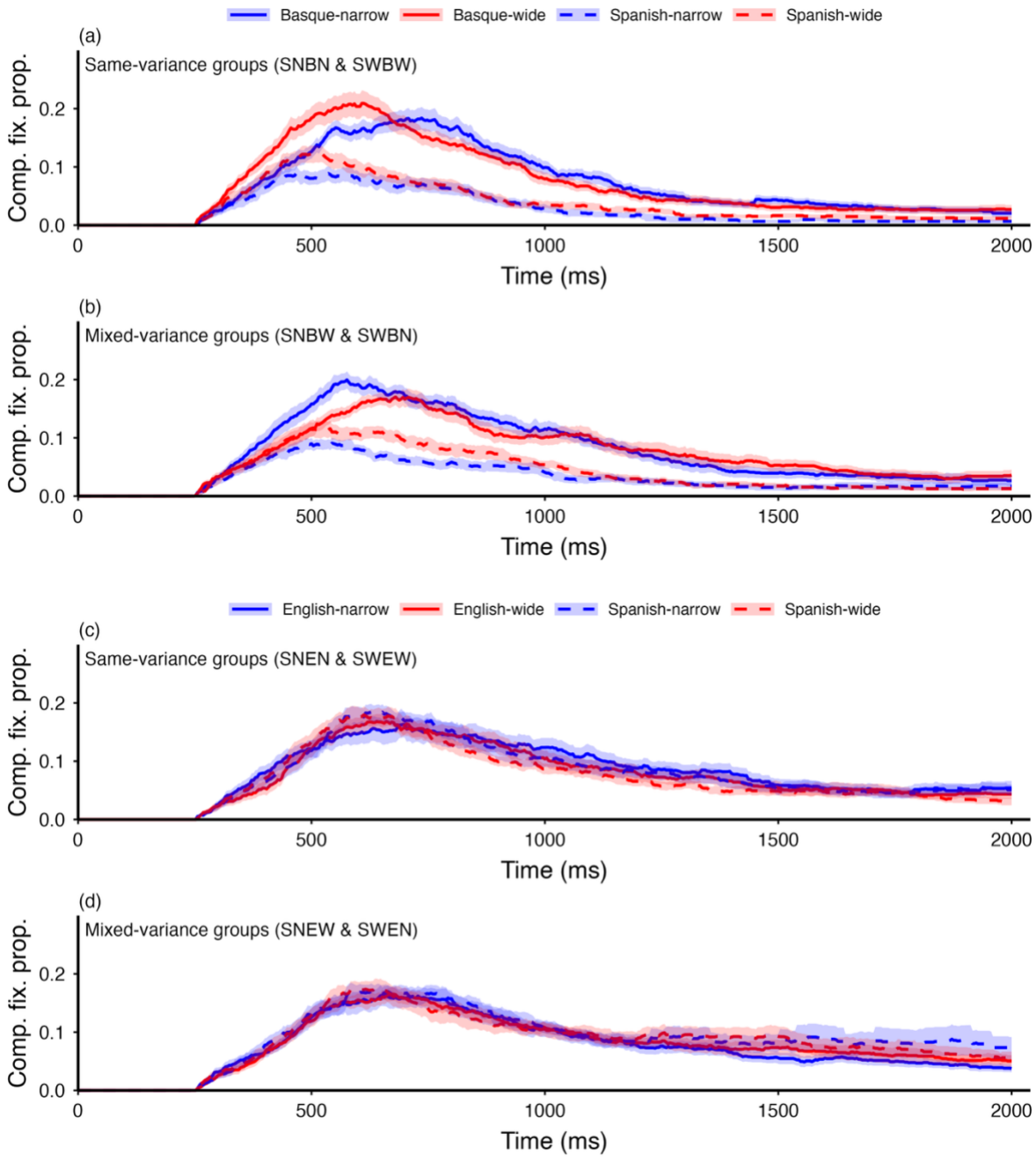
Taken together, these results suggest that the effect of distributional variance on categorization slope was present regardless of whether the distributional variance differed across languages, but that the magnitude of the effect was relatively small.

Next, we focus on a more sensitive measure reported by Clayards et al. (2008): the proportion of fixations on the competitor picture. We first examined the same-variance groups (i.e., SNBN and SWBW). Figure *2.6*a shows the proportion of fixations on the competitor pictures in the wide and narrow conditions for these two groups. Recall that only

the ambiguous stimuli (i.e., −15 to +15 ms rVOT) and trials in which participants ultimately clicked on the target were included in this analysis. If the effect of distributional learning on gradiency generalizes to a bilingual context, listeners who were exposed to the wide condition (red curves) should look more toward the competitor than those exposed to the narrow condition (blue curves). Indeed, Figure *2.6*a shows this pattern, consistent with the descriptive statistics (Basque-narrow: M = 6.9%, SD = 8.5%; Basque-wide: M = 7.2%, SD = 5.2%; Spanish-narrow: M = 2.0%, SD = 3.6%; Spanish-wide: M = 3.8%, SD = 4.2%). Statistically, we found decisive evidence for an effect of distributional variance, such that listeners fixated more on the competitor in the wide condition, β = 0.40, 95% *CrI* [0.13, 0.66], BF = 149.12, suggesting that gradiency is malleable based on distributional statistics. We also found strong evidence for a main effect of language, with more fixations in Basque than in Spanish, β = 0.49, 95% *CrI* [0.15, 0.79], BF = 23.10. Evidence for an interaction between distributional variance and language weakly favored the null hypothesis, β = 0.12, 95% *CrI* [−0.20, 0.45], BF = 0.43. Full results are presented in Table A2.4.

Figure 2.6

Proportion of Fixations on Competitor Pictures by Distributional Variance and Language in (a) the SNBN and SWBW Groups, (b) the SNBW and SWBN Groups, (c) the SNEN and SWEW Groups, and (d) the SNEW and SWEN Groups



Note. The y-axis represents competitor fixation proportion. The shaded ribbons around each curve represent the standard error of the mean (±1 *SE*) across participants at each time point. Fixation proportions during the first 250 ms after stimulus onset were set to zero, corresponding to the 50 ms onset silence and the typical 200 ms oculomotor delay, as this period does not reflect lexical processing. Only trials with ambiguous stimuli (rVOT = −15 to +15 ms) in which participants ultimately clicked on the target were included.

Figure *2.6*b shows fixation patterns as a function of distributional variance and language for the SNBW and SWBN groups. Although the graph appears to show a reversal effect for Basque, with a higher peak in the narrow than in the wide condition, the descriptive statistics within the analysis time window (250–2,000 ms) indicate similar average proportions of competitor fixations between conditions for Basque (Basque–narrow: M = 7.0%, SD = 7.8%; Basque–wide: M = 6.9%, SD = 6.6%; Spanish–narrow: M = 3.5%, SD = 7.4%; Spanish–wide: M = 4.2%, SD = 5.0%). There was moderate-to-strong evidence that listeners fixated more on the competitor in the wide than in the narrow condition, β = 0.29, 95% *CrI* [0.01, 0.56], BF = 10.78, suggesting that listeners were more gradient when the distributional variance was wider. There was also moderate evidence for a main effect of language, with more competitor fixations in Basque than in Spanish, β = 0.37, 95% *CrI* [0.07, 0.65], BF = 9.59. Evidence weakly favored the null hypothesis for the interaction between distributional variance and language, β = 0.01, 95% *CrI* [−0.38, 0.40], BF = 0.36. Full results are presented in Table A2.5. Based on the fixation results, gradiency appears to be malleable regardless of whether the two languages had the same or different distributional variance.

To ensure that participants across the four groups were comparable on relevant background characteristics that could influence categorization slopes and fixation patterns, we compared Spanish and Basque self-rated listening proficiency and exposure, language-switching frequency, musical training duration, and Spanish and Basque LexTale scores across the theoretically relevant between-condition contrasts (i.e., SNBN vs. SWBW and SNBW vs. SWBN). None of these variables showed consistent or strong evidence for group differences. Bayes factors from the Bayesian independent-samples t-tests were generally small (BF range = 0.32–1.60), indicating either evidence favoring the null hypothesis or, at most, weak evidence for differences. Consequently, these background factors are unlikely to

account for the observed differences in categorization slopes and fixation patterns across conditions.

2.4 Interim Discussion

In summary, we observed effects in the expected direction for both measures: categorization slope of click response curves and competitor fixation proportion. Although the evidence for the click response measure was weak to moderate, competitor fixation provides a more sensitive measure and is also the measure previously shown to be associated with VAS gradiency (Kapnoula & McMurray, 2021). Taken together, these findings indicate that gradiency is malleable in bilingual contexts. Specifically, exposure to wider distributions led to enhanced gradiency. Crucially, despite the apparent reversal of the effect for Basque (see Figure *2.6*b), the analyses showed that the distributional effects on gradiency did not depend on whether the two languages differed in their VOT distributions, suggesting that listeners can track and adapt to distinct distributional statistics across languages within the same individual, even when both languages are produced by the same talker. This suggests that gradiency was enhanced following exposure to a wide distribution, regardless of whether distributional variance was the same or different across the two languages.

Although this was not the central focus of the study, we observed a main effect of language, such that participants fixated more on competitor pictures in Basque than in Spanish. This difference may partly reflect variation in language proficiency. Although participants were highly proficient in both languages, their Basque LexTale scores (M = 91.0, SD = 6.9) were slightly lower than their Spanish scores (M = 93.9, SD = 4.7; see similar patterns in picture-naming performance in Table A2.1). Another plausible explanation concerns asymmetries in bilinguals' language-switching patterns in daily life. Spanish–Basque bilinguals may switch from Basque to Spanish more frequently than from Spanish to Basque (Ibarra Murillo, 2013; Lantto, 2015), which could result in stronger sustained

activation of Spanish even during Basque processing. Within the context of the present study, such asymmetric activation may increase lexical competition or uncertainty in Basque. Nonetheless, this interpretation remains speculative, as we did not collect information about the direction of language switching. Moreover, other factors cannot be ruled out, including differences in lexical frequency or the visual salience of the stimuli. This possibility is particularly relevant given the limited availability of Basque–Spanish item pairs that met the experimental criteria.

We next examine whether the malleability of gradiency extends to language pairs with more divergent VOT distributions. Whereas the VOT ranges of Spanish and Basque largely overlap (Souganidis et al., 2024), English exhibits substantially longer VOTs than both Spanish (Allen & Miller, 1999) and Basque, given the equivalence of Spanish and Basque VOT distributions. This contrast allows us to test whether distributional learning effects on gradiency generalize across languages with markedly different phonetic implementations of voicing. One possibility is that distributional learning becomes more challenging when the VOT distributions of two languages differ markedly (e.g., in the Spanish–English case compared with the Spanish–Basque case), because listeners cannot readily apply their L1-based priors to interpret L2 input (Pajak et al., 2016). This may result in a weaker distributional learning effect. Alternatively, the malleability of gradiency driven by distributional learning may instead be stronger when the two languages differ more in their VOT distributions, because less overlapping ranges could make it easier for listeners to form and maintain distinct language-specific models. It is also possible that cross-linguistic overlap in VOT does not influence the distributional learning effect. To disentangle these possibilities, we conducted Experiment 1.2 with Spanish–English bilinguals. By combining these data with those from Experiment 1.1, we tested whether cross-linguistic differences in the degree of VOT overlap modulate the malleability of gradiency.

2.5 Method of Experiment 1.2

2.5.1 Participants

We planned to recruit 72 participants, as in Experiment 1.1, with half drawn from the University of Connecticut (UConn) community and half from BCBL. At UConn, participants were required to meet several inclusion criteria: they had to self-rate their Spanish proficiency as at least 6 out of 10, feel comfortable speaking both English and Spanish, and be between 18 and 35 years of age. At BCBL, participants demonstrated high proficiency in both Spanish and English by scoring at least 85% on a picture-naming task. In total, 85 participants were recruited, as 13 participants were excluded and replaced for the following reasons: Three participants did not consistently fixate on the target picture when clicking on it (i.e., fixated on fewer than 60% of trials within the 300-ms window preceding the click). Five participants were excluded because their /p/-response curve never crossed the 0.5 proportion threshold (i.e., they consistently selected the /b/-onset picture even when the stimulus was a clear /p/) in either Spanish or English. Four participants were excluded due to technical issues, and one participant did not complete the experiment. The final sample consisted of 72 bilinguals (52 females; 20 males) aged between 18 and 35 years (M = 22.6, SD = 4.5): 36 English–Spanish bilinguals tested at UConn and 36 English–Spanish bilinguals tested at BCBL.

All participants reported no history of hearing, reading, motor, or speech disorders, and no neurological conditions or head injuries. All had normal or corrected-to-normal vision and no color blindness. Participants reported a wide range of Spanish and English dialectal backgrounds. For Spanish, the most commonly reported dialect was Peninsular Spanish (N = 36), followed by Colombian and Mexican Spanish (N = 8 each). Smaller numbers of participants reported Peruvian (N = 6), Caribbean Spanish (including Dominican and Puerto

Rican; $N = 7$), Ecuadorian ($N = 3$), Central American ($N = 2$), and Rioplatense ($N = 1$), with one participant indicating that they did not know their Spanish dialect. For English, most participants reported American English ($N = 39$), followed by British English ($N = 9$), a mixed British–American variety ($N = 6$), and Canadian English ($N = 1$). Some participants reported not knowing their English dialect ($N = 17$). Table A2.6 shows participant characteristics of Experiment 1.2. The study was approved by the University of Connecticut Institutional Review Board (protocol #X24-0247) and by the BCBL Ethics Review Board (ref. 080625BC), in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants prior to participation. Participants received either course credit or monetary compensation for their time.

2.5.2 Task Overview

The design was identical to that of Experiment 1.1, except that all instances of *Basque* were replaced with *English*. In other words, the four experimental conditions were Spanish-narrow/English-narrow (SNEN), Spanish-wide/English-wide (SWEW), Spanish-narrow/English-wide (SNEW), and Spanish-wide/English-narrow (SWEN).

2.5.3 Stimuli

The stimulus design was identical to that of Experiment 1.1 (see Section 2.2.3). Due to the need to match the number of syllables in the English stimuli, some of the Spanish stimuli differed from those used in Experiment 1.1. No Spanish items that vary across Peninsular and Latin American dialects were included. All stimuli contained two syllables and followed the same stress pattern. The Spanish and English stimuli are listed in Table 2.3.

Table 2.3
Stimulus Sets Used in Experiment 1.2

Set	Language	Critical minimal pair	English translation	Filler	English translation
1	Spanish	bata – pata	robe – paw	lana – rana	wool – frog
2	Spanish	viña – piña	vineyard – pineapple	loncha – roncha	slice – welt
3	Spanish	baño – paño	bathroom – cloth	lamo – ramo	I lick – bouquet
4	English	butter – putter		lighter – writer	
5	English	barking – parking		locker – rocker	
6	English	boring – pouring		locket – rocket	

Note. The first stimulus set had been used in Experiment 1.1.

The recording procedure was identical to that of Experiment 1.1, unless otherwise specified. All words were recorded by a female Spanish–English bilingual speaker affiliated with UConn, who spoke Spanish with a Latin American (Mexican) accent and English with an American accent. [3] Each word was produced within a carrier phrase: *Es mi* ___ for Spanish and *It's my* ___ for English, ensuring consistent prosody across items. To obtain prevoiced VOTs for English /b/ tokens and long-lag VOTs for Spanish /p/ tokens for manipulation, the speaker was asked to produce English /b/ as in Spanish and Spanish /p/ as in English.

The VOTs of the critical stimuli (/b/ and /p/) were manipulated to create 12 10-ms steps using the method introduced in Section 2.2.3. For Spanish, the VOT ranged from −65 to +45 ms (as in Experiment 1.1), and for English, it ranged from −30 to +80 ms (following Clayards et al., 2008). The number of trials at each VOT value for the narrow and wide

[3] Originally, we planned to recruit all 72 participants at UConn; therefore, all stimuli were recorded with accents that are mainstream in that region: a Latin American (Mexican) accent for Spanish and an American accent for English. However, due to difficulties in recruiting participants at UConn, we ultimately decided to recruit 36 participants in San Sebastian, where many speakers are more familiar with Peninsular Spanish and some may be familiar with British-accented English. Despite this difference, the same recordings were used for all participants, given established evidence of similar VOT values for bilabial contrasts across Latin American and Peninsular Spanish dialects (e.g., Lisker & Abramson, 1964; Rosner, 2000; Souganidis et al., 2024), as well as across American and British English dialects (e.g., Chao & Chen, 2008; Docherty, 1992; Lisker & Abramson, 1964). Relatedly, each participant's category boundary was taken into account in the analyses.

variance conditions in English is shown in Table 2.4. The F_0 was fixed at 190 Hz across all stimuli.

Table 2.4
Number of Trials per VOT Value in the Narrow and Wide Variance Conditions for English

VOT (ms)	−30	−20	−10	0	10	20	30	40	50	60	70	80
Narrow	0	3	27	54	27	3	3	27	54	27	3	0
Wide	3	12	27	30	27	15	15	27	30	27	12	3

Note. Trial numbers and VOT values are identical to those used in Clayards et al. (2008).

2.5.4 Procedure

The procedure of the eye-tracking experiment was identical to Experiment 1.1; however, the questionnaire differed in several respects. First, participants at UConn completed both the English (Lemhöfer & Broersma, 2012) and Spanish (Izura et al., 2014) versions of the LexTALE to assess their language proficiency. The LexTALE tests were administered in a questionnaire format at the end of the session and were identical to those completed by participants at BCBL during registration. In each version, participants indicated whether each item was a real word or a nonword. The Spanish LexTALE contained 90 items, and the English version contained 60 items. In addition, participants at UConn answered several additional demographic and language-background questions, including their age of acquisition (AoA) for Spanish and English, as well as their age, gender, ethnicity, and race. Participants tested at BCBL did not complete the LexTALE or the additional questions described above, as this information had already been collected during registration. Moreover, all participants provided information about their English and Spanish dialects. The entire session lasted approximately 90 minutes for UConn and 60 minutes for BCBL.

2.5.5 Eye-tracking Recording and Analyses

All procedures were identical to those described in Section 2.2.5 for the BCBL participants. For the UConn participants, the only differences were that (1) an EyeLink 1000 (rather than an EyeLink 1000 Plus) was used, (2) the headphone set was a Sennheiser HD 595, and (3) the experiment was run using Experiment Builder version 2.6.11.

2.5.6 Statistical Analyses

Before conducting data analyses, all filler trials (32,832 out of 65,664; 50% of total trials) were excluded. Next, 606 experimental trials (1.9% of experimental trials) with RTs exceeding ±3 standard deviations from each participant's mean RT were classified as outliers and removed. An additional 71 trials (0.2% of experimental trials) containing responses to filler items were also excluded. Finally, 2,207 trials (6.7% of experimental trials) were removed based on curve-fitting quality, as they belonged to participant–item combinations with response curves showing poor fits ($R^2 < .80$). This resulted in a total of 29,947 trials remaining for analysis.

As in Experiment 1.1, the two dependent variables were the categorization slope and the proportion of fixations on competitor pictures. The slope parameter was obtained using the same curve-fitting procedure described in Section 2.2.7. Model fits for the remaining data were excellent (average $R^2 = .982$).

For the fixation analyses, we again created a relative VOT measure (rVOT), calculated with respect to the category boundary estimated from the logistic function fitted to each participant's mouse-click responses. A total of 2,013 trials (6.1% of experimental trials) in which participants selected the competitor (i.e., clicking on /p/-onset pictures for negative rVOT values or on /b/-onset pictures for positive rVOT values) were excluded, leaving 27,934 trials for analysis. For the final analyses, we focused on trials with rVOT values

between −15 and +15 ms, which correspond to the most ambiguous region of the continuum. These trials constituted 22.6% of the total experimental trials across both research questions.

We first repeated the statistical analyses reported in Section 2.2.7, with the only difference being that English replaced Basque. We then combined the datasets from both experiments to compare the distributional learning effect across language-pair contexts, focusing separately on same-variance and mixed-variance groups. For both the categorization slope and competitor-fixation measures, we used Bayesian linear mixed-effects models with distributional variance (narrow vs. wide) and experiment (Spanish–Basque vs. Spanish–English) as fixed effects, along with their interaction. Evidence for an experiment × distributional variance interaction would indicate that differences in VOT overlap across language pairs modulate the strength of distributional learning.

2.6 Results of Experiment 1.2

As in Experiment 1.1, we first verified that participants' overall response patterns were consistent with expectations. Figure *2.3*b shows both individual and average mouse-click responses across VOT values for Spanish–English bilinguals. As shown in the figure, all participants exhibited an increasing proportion of clicks on /p/-onset pictures as VOT increased. Importantly, the average category boundary (defined as the VOT value corresponding to a 0.5 proportion of /p/ responses) differed between Spanish and English in a manner consistent with prior literature, with the English boundary occurring at a higher VOT than the Spanish boundary. This pattern indicates that, on average, participants maintained distinct VOT representations for the two languages. A comparison of Figure *2.3*a and 2.3b further suggests that Spanish–English bilinguals exhibited more gradient response patterns than Spanish–Basque bilinguals. Although no explicit time limit was imposed, participants typically responded quickly (Spanish: M = 1,621 ms, SD = 551 ms; English: M = 1,622 ms, SD = 521 ms).

Next, we examined overall fixation patterns to the target, competitor, and filler pictures. Figure *2.4*b shows both individual and grand-average fixation trajectories over time. Fixations on the target increased as the trial unfolded, whereas fixations on the competitor and filler pictures showed an early, transient peak. In contrast to Experiment 1.1 (Figure *2.4*a), the present results show the expected pattern, with competitor fixations visually higher than filler fixations.

We now turn to the results addressing our first research question: whether Spanish–English bilingual listeners were more gradient when exposed to the wide-variance distribution. We addressed this question by examining the same-variance groups (i.e., SNEN and SWEW). The left panel of Figure *2.5*b shows the corresponding participant-by-item categorization slope values for the narrow and wide distributional-variance conditions in Spanish and English (English–narrow: $M = 0.094$, $SD = 0.013$; English–wide: $M = 0.069$, $SD = 0.032$; Spanish–narrow: $M = 0.090$, $SD = 0.022$; Spanish–wide: $M = 0.070$, $SD = 0.031$). Bayesian mixed-effects modeling revealed decisive evidence for a distributional learning effect, such that categorization slopes were shallower in the wide than in the narrow condition, $\beta = -0.60$, 95% *CrI* [−0.91, −0.28], $BF = 38{,}662.81$. This contrasts with the moderate evidence observed in Experiment 1.1 and indicates that the distributional learning effect was more robust. In addition, there was weak evidence supporting the null hypothesis for the main effect of language, $\beta = 0.14$, 95% CrI [−0.25, 0.51], $BF = 0.38$, suggesting that categorization slopes did not reliably differ between Spanish and English. There was inconclusive evidence for an interaction between distributional variance and language, $\beta = -0.19$, 95% *CrI* [−0.62, 0.22], $BF = 0.92$. Full results are reported in Table A2.7.

We then turned to our second research question, examining whether gradiency is malleable within individuals across languages, focusing on data from the mixed-variance groups (i.e., SNEW and SWEN). The right panel of Figure *2.5*b shows the corresponding

participant-by-item categorization slope values for the narrow and wide distributional-variance conditions in Spanish and English. (English–narrow: $M = 0.092$, $SD = 0.019$; English–wide: $M = 0.078$, $SD = 0.029$; Spanish–narrow: $M = 0.085$, $SD = 0.028$; Spanish–wide: $M = 0.069$, $SD = 0.032$). In contrast to the weak evidence reported in Experiment 1.1, Bayesian mixed-effects modeling revealed decisive evidence for a distributional learning effect, such that categorization slopes were shallower in the wide than in the narrow condition, $\beta = -0.46$, 95% *CrI* [−0.77, −0.14], $BF = 132.69$. This indicates that gradiency is malleable within individuals when distributional variance differs across the two languages. In addition, there was inconclusive evidence regarding the main effect of language, $\beta = 0.22$, 95% *CrI* [−0.22, 0.61], $BF = 1.20$. There was also weak evidence in favor of the null hypothesis for the interaction between distributional variance and language, $\beta = 0.01$, 95% *CrI* [−0.41, 0.41], $BF = 0.52$. Full model results are reported in Table A2.8.

We then analyzed the proportion of fixations on the competitor for ambiguous stimuli (−15 to +15 ms rVOT) on target-clicking trials, focusing on the same-variance groups (SNEN and SWEW). Figure *2.6*c shows competitor fixation proportions in the wide and narrow conditions for both Spanish and English in these two groups. If gradiency is malleable through exposure (as observed for the same measure in Experiment 1.1), participants in the wide condition (red curve) would be expected to fixate on the competitor more than those in the narrow condition (blue curve). Instead, the two curves largely overlap (English–narrow: $M = 9.1\%$, $SD = 10.2\%$; English–wide: $M = 7.3\%$, $SD = 7.1\%$; Spanish–narrow: $M = 8.7\%$, $SD = 9.4\%$; Spanish–wide: $M = 7.2\%$, $SD = 6.8\%$). The Bayesian analyses provided moderate evidence for a null effect of distributional variance, $\beta = 0.00$, 95% *CrI* [−0.28, 0.29], $BF = 0.23$, indicating no reliable difference in gradiency between the two distributional conditions. There was also moderate evidence for a null effect of language, $\beta = 0.03$, 95% *CrI* [−0.25, 0.31], $BF = 0.20$. Evidence for an interaction between distributional variance and

language weakly favored the null hypothesis, $\beta = -0.07$, 95% *CrI* [−0.45, 0.33], *BF* = 0.40. Full results are reported in Table A2.9.

Figure *2.6*d shows the proportion of fixations on the competitor in the wide and narrow conditions for both Spanish and English in the mixed-variance groups (i.e., SNEW and SWEN). Descriptively, competitor fixation proportions were similar across distributional conditions and languages (English–narrow: *M* = 7.4%, *SD* = 9.0%; English–wide: *M* = 8.1%, *SD* = 6.9%; Spanish–narrow: *M* = 8.7%, *SD* = 9.2%; Spanish–wide: *M* = 8.6%, *SD* = 8.4%). As in the previous analysis, listeners showed broadly comparable competitor fixations across conditions. Bayesian analyses provided weak evidence in favor of more competitor fixations in the wide condition, $\beta = 0.16$, 95% *CrI* [−0.15, 0.47], *BF* = 2.55. The 95% credible interval included zero, indicating that the effect was not robust. We also found weak evidence in favor of a null effect of language, $\beta = -0.16$, 95% *CrI* [−0.56, 0.26], *BF* = 0.38. Evidence for the interaction between distributional variance and language weakly favored the null hypothesis, $\beta = 0.16$, 95% *CrI* [−0.28, 0.59], *BF* = 0.61. Full results are presented in Table A2.10.

To ensure that participants in each group were comparable on relevant characteristics that could influence categorization slopes and fixation patterns, we compared Spanish and English self-rated listening proficiency and exposure, language-switching frequency, musical training duration, and Spanish and English LexTALE scores across the theoretically relevant between-condition contrasts (i.e., SNEN vs. SWEW and SNEW vs. SWEN). All Bayesian independent-samples t-tests favored the null hypothesis (*BF*s ranged from 0.32 to 0.57). These results indicate that participants were well matched across conditions with respect to linguistic experience, proficiency, and related individual-difference measures. Accordingly, these variables are unlikely to have contributed to the observed differences in the categorization slopes and fixation patterns.

2.7 Results of the Combined Analysis

Here we report analyses based on the combined dataset from Experiments 1.1 and 1.2 to address our final research question: whether the degree of cross-linguistic VOT overlap modulates the malleability of gradiency. Table *2.5* summarizes the Bayesian evidence for the distributional learning effect across experiments and conditions.

Table 2.5
Bayesian Evidence for the Distributional Learning Effect

	SNBN and SWBW	SNBW and SWBN	SNEN and SWEW	SNEW and SWEN
Categorization slope	Moderate (*BF* = 3.92)	Weak/inconclusive (*BF* = 1.41)	Decisive (*BF* = 38,662.81)	Strong (*BF* = 132.69)
Competitor fixation	Decisive (*BF* = 149.12)	Moderate to strong (*BF* = 10.78)	Moderate for the null (*BF* = 0.23)	Weak (*BF* = 2.55)

We first compared categorization slopes in the same-variance groups (Spanish–Basque: SNBN vs. SWBW; Spanish–English: SNEN vs. SWEW), collapsed across language. Descriptively, in both experiments, slopes were shallower in the wide than in the narrow condition (Spanish–Basque: narrow: $M = 0.099$, $SD = 0.01$; wide: $M = 0.095$, $SD = 0.01$; Spanish–English: narrow: $M = 0.092$, $SD = 0.02$; wide: $M = 0.070$, $SD = 0.03$). Bayesian mixed-effects modeling revealed decisive evidence for a main effect of distributional variance, such that categorization slopes were shallower in the wide than in the narrow condition, $\beta = -0.20$, 95% *CrI* [−0.43, 0.03], $BF = 30{,}649.45$. This result supports the presence of the distributional learning effect in a bilingual context. There was also decisive evidence for a main effect of experiment, reflecting overall shallower slopes in Spanish–English than in Spanish–Basque bilinguals, $\beta = -0.34$, 95% *CrI* [−0.58, −0.10], $BF = 5{,}436.65$. Critically, the experiment × distributional variance interaction was also supported by decisive evidence, $\beta = -0.70$, 95% *CrI* [−1.02, −0.38], $BF = 2{,}148.70$, indicating that the

distributional learning effect was stronger in Spanish–English than in Spanish–Basque bilinguals. Table A2.11 shows complete results.

We then turned to the mixed-variance groups. Descriptively, categorization slopes were shallower in the wide than in the narrow condition in both experiments (Spanish–Basque: narrow: $M = 0.100$, $SD = 0.003$; wide: $M = 0.097$, $SD = 0.010$; Spanish–English: narrow: $M = 0.088$, $SD = 0.020$; wide: $M = 0.074$, $SD = 0.030$). Bayesian mixed-effects analyses provided decisive evidence for a main effect of distributional variance, such that categorization slopes were shallower in the wide than in the narrow condition, $\beta = -0.14$, 95% *CrI* [−0.36, 0.09], $BF = 545.57$. These results provide clear evidence for the presence of a distributional learning effect in bilingual listeners. There was also decisive evidence for a main effect of experiment, reflecting overall shallower slopes in Spanish–English than in Spanish–Basque bilinguals, $\beta = -0.59$, 95% *CrI* [−0.89, −0.32], $BF = \infty$. Importantly, the experiment × distributional variance interaction was supported by strong evidence, $\beta = -0.49$, 95% *CrI* [−0.81, −0.18], $BF = 60.26$, indicating a larger distributional learning effect in Spanish–English than in Spanish–Basque bilinguals, consistent with the pattern observed in the same-variance groups. Table A2.12 shows complete results.

We next report analyses of competitor fixation proportions for the same-variance groups, collapsed across language. Descriptively, competitor fixations were slightly higher in the wide than in the narrow condition in Spanish–Basque bilinguals (narrow: $M = 4.4\%$, $SD = 6.9\%$; wide: $M = 5.4\%$, $SD = 5.0\%$), whereas the opposite pattern was observed in Spanish–English bilinguals (narrow: $M = 8.9\%$, $SD = 9.8\%$; wide: $M = 7.2\%$, $SD = 6.9\%$). Bayesian mixed-effects modeling provided weak evidence for a main effect of distributional variance, $\beta = 0.40$, 95% *CrI* [0.19, 0.60], $BF = 3.47$, such that competitor fixations were higher in the wide than in the narrow condition. Overall, this indicates that the distributional learning effect was weak across experiments. There was decisive evidence for a main effect of

experiment, $\beta = 0.57$, 95% *CrI* [0.32, 0.83], *BF* = 109.42, reflecting overall more competitor fixations in Spanish–English than in Spanish–Basque bilinguals. Crucially, there was moderate-to-strong evidence for an experiment × distributional variance interaction, $\beta = -0.39$, 95% *CrI* [−0.68, −0.11], *BF* = 10.54, indicating that the effect of distributional variance on competitor fixation differed across experiments. In contrast to the categorization slope measure, gradiency, as indexed by fixations on the competitor, appears to be more malleable in Spanish–Basque than in Spanish–English bilinguals for the same-variance groups. Full results are reported in Table A2.13.

For the mixed-variance groups, the proportion of competitor fixations was slightly higher in the wide than in the narrow condition in both the Spanish–Basque experiment (narrow: *M* = 5.2%, *SD* = 7.8%; wide: *M* = 5.6%, *SD* = 6.0%) and the Spanish–English experiment (narrow: *M* = 8.0%, *SD* = 9.1%; wide: *M* = 8.3%, *SD* = 7.7%). Bayesian analyses provided decisive evidence for a main effect of distributional variance, $\beta = 0.28$, 95% *CrI* [0.10, 0.45], *BF* = 218.38, indicating more competitor fixations in the wide condition. This result provides clear evidence that gradiency is malleable when listeners are exposed to different levels of distributional variance. There was also decisive evidence for a main effect of experiment, $\beta = 0.50$, 95% *CrI* [0.27, 0.76], *BF* = 27,020.35, reflecting overall more competitor fixations in Spanish–English than in Spanish–Basque bilinguals. In contrast to the same-variance analysis, there was moderate evidence in favor of the null hypothesis for the experiment × distributional variance interaction, $\beta = 0.00$, 95% *CrI* [−0.25, 0.25], *BF* = 0.25, suggesting that the magnitude of the distributional learning effect did not reliably differ between Spanish–Basque and Spanish–English bilinguals in the mixed-variance groups. Full results are reported in Table A2.14.

2.8 Exploratory Analysis

By inspection of Table *2.5*, the distributional variance effect appears numerically smaller in the mixed-variance groups than in the same-variance groups across most comparisons (except for the competitor fixation measure in Experiment 1.2). To test this pattern directly, we collapsed the data across experiments and fitted Bayesian mixed-effects models with distributional variance (narrow vs. wide), group (same-variance vs. mixed-variance), and their interaction as predictors. A distributional variance × group interaction would indicate that the magnitude of the distributional learning effect depends on whether the two languages share the same distributional variance or differ in the distributional variance of the stimuli presented during the experiment. For categorization slope, evidence for the interaction was inconclusive, $\beta = 0.01$, 95% *CrI* [−0.56, 0.52], $BF = 1.00$ (see Table A2.15), indicating no reliable evidence for either the presence or absence of an interaction. For competitor fixation, there was moderate evidence in favor of the null hypothesis for the interaction, $\beta = 0.07$, 95% *CrI* [−0.11, 0.24], $BF = 0.22$ (see Table A2.16). This suggests that although the distributional learning effect was numerically smaller in the mixed-variance groups, this reduction was not statistically credible. Taken together, these results indicate that distributional learning is robust in a bilingual context and does not depend on whether the two languages share or differ in distributional variance.

2.9 Discussion

Across two experiments, the present study examined (1) whether gradiency is malleable in a bilingual context, (2) whether bilinguals can simultaneously maintain distinct probability distributions of acoustic-phonetic cues for two languages spoken by the same talker, and how this information in turn shapes gradiency, and (3) whether the degree of cross-linguistic VOT overlap modulates the malleability of gradiency. Across measures, the results indicate that gradiency is malleable in bilingual listeners, with the exception of the

competitor-fixation measure in Spanish–English bilinguals. However, the answers to the latter two questions were less straightforward and appeared to depend on the measure used. In the mixed-variance groups, distributional learning effects were evident in Spanish–Basque bilinguals when indexed by competitor fixation, but the effects were weak or inconclusive when assessed using the categorization slope measure. In contrast, for Spanish–English bilinguals, the distributional learning effect was evident in the mixed-variance groups when assessed using the categorization slope measure, but was weak when indexed by competitor fixation. Similarly, the role of cross-linguistic VOT overlap differed across measures. Greater VOT overlap between Spanish and Basque appeared to attenuate distributional learning effects when indexed by categorization slope, but boosted them in the same-variance groups when indexed by competitor fixation (though not in the mixed-variance groups). Taken together, while we assume that both categorization slope and competitor fixation can be interpreted as behavioral indices of gradiency in distributional learning, the present findings suggest that these measures may reflect partially distinct aspects of the learning process rather than gradiency alone.

It is important to note that Clayards et al. (2008) is, to our knowledge, the only study to have examined distributional learning using both categorization slope and competitor fixation measures, and it reported converging evidence across the two indices. Subsequent studies have typically relied on only one measure, using either competitor fixation (e.g., Nixon et al., 2016) or categorization slope (e.g., Gür & McMurray, 2025; Theodore & Monto, 2019; Theodore et al., 2020). Moreover, the analytical methods used by Clayards et al. (2008) differ from ours in at least two respects. First, they focused on raw VOT and did not take participants' category boundaries into account when deciding ambiguous sounds. Second, they included all trials in the analysis, including those in which participants clicked on the competitor, making it less clear what the measure captured. We therefore believe that

our measure more directly taps into gradiency. Taken together, it remains possible that categorization slope and competitor fixation capture partially distinct aspects of distributional learning.

In particular, distributional learning may involve at least two dimensions: gradiency and consistency. As discussed below, previous studies have yielded mixed findings, suggesting that categorization slope and competitor fixation may differentially index these two dimensions; however, no definitive conclusion has been reached regarding which measure corresponds to which dimension. To preview our interpretation, we propose that categorization slope is more closely related to consistency, whereas competitor fixation is more closely related to gradiency.

Regarding the categorization slope measure, Honda et al. (2024a) showed that the slope derived from 2AFC tasks is more strongly associated with VAS consistency than with gradiency per se, with a steeper slope reflecting more consistent responses. In the present eye-tracking task, excluding filler stimuli, participants effectively performed a 2AFC decision. Combining the findings by Honda et al. (2024a) with this task structure, the slope measure in the present study may therefore primarily reflect consistency. Therefore, we propose that a higher categorization slope may primarily reflect increased consistency rather than enhanced gradiency.

This proposal is also supported by one of our findings: Spanish–English bilinguals exhibited shallower categorization slopes than Spanish–Basque bilinguals (Figure *2.3*). Spanish–English bilinguals showed, on average, lower proficiency in both languages than Spanish–Basque bilinguals (mean LexTALE scores of around 80 vs. around 90), despite efforts to match the groups. Lower proficiency may increase uncertainty and therefore lead to greater inconsistency (rather than higher gradiency) overall.

Another possible explanation for the overall shallower categorization slopes observed in Spanish–English bilinguals compared with Spanish–Basque bilinguals concerns the difference in the combined VOT range across the two languages. Although identical within-language VOT steps were used across experiments, the overall VOT range differed substantially, with the Spanish–English experiment spanning a wider range (−65 to 80 ms) than the Spanish–Basque experiment (−65 to 45 ms). In distributional learning paradigms, wider distributional variance is known to produce shallower categorization slopes, and such variance is often implemented by a broader stimulus range. Notably, in the present study, the between-experiment difference in VOT range is substantially larger than the wide-versus-narrow distributional manipulation implemented within each experiment. From this perspective, the shallower slopes observed in the Spanish–English experiment may reflect a mechanism similar to that underlying distributional learning. Consistent with the interpretation above, this pattern suggests that listeners became more uncertain and therefore less consistent when faced with a wide range of VOT values, whether arising from the experimental stimulus distribution or from long-term cross-language exposure.

Furthermore, we did not only observe overall shallower categorization slopes in Spanish–English bilinguals compared with Spanish–Basque bilinguals, but also a stronger effect of distributional variance in the former group, reflected in particularly shallow slopes in the wide condition. This pattern can be explained by the two factors discussed above. Specifically, lower language proficiency together with a wider overall VOT range may increase perceptual uncertainty, thereby increasing inconsistency in categorization.

The competitor-fixation measure, which is often considered more sensitive than the categorization-slope measure based on explicit mouse-click responses, appears to face a similar interpretive challenge. Previous work has shown that competitor fixation can index

gradiency (Kapnoula & McMurray, 2021). As discussed earlier, Kapnoula and McMurray (2021) demonstrated that listeners who were more gradient, as measured by a VAS, exhibited more graded patterns of competitor fixation across the VOT continuum on trials in which they ultimately selected the target. From this perspective, competitor fixation in the VWP may likewise reflect gradiency. At the same time, it is important to note that the VWP may index a different aspect of gradiency than that captured by the VAS. VAS-derived gradiency is well suited for capturing warping around the category boundary, whereas VWP-based measures place greater emphasis on within-category sensitivity, as fixation analyses typically bin stimuli by category before examining competitor activation (Kapnoula, 2016). In addition, compared with the VAS, the VWP requires a large number of trials, which may alter listeners' processing strategies through repeated exposure (Apfelbaum et al., 2022). Moreover, because eye fixations are inherently binary at any given moment, evidence for gradient speech processing in the VWP relies on averaging across time and trials, making it difficult to rule out alternative interpretations, such as rapid vacillation between discrete perceptual states (Apfelbaum et al., 2022). Future studies should therefore make greater use of the VAS to more directly assess the malleability of gradiency.

However, competitor fixation may also be related to consistency. This issue is tied to the question of whether distributional learning effects indexed by competitor fixation reflect a learning mechanism per se. As noted earlier, Nixon et al. (2016) argued that competitor fixation does not necessarily index cumulative perceptual learning, but may instead reflect a verification process in which increased auditory uncertainty leads listeners to seek additional visual information to confirm their decisions. Consistent with this interpretation, the present data also suggest that increased fixations on the competitor do not reflect cumulative learning over time. If competitor fixation indexed learning, one would expect an interaction between distributional variance and trial number for ambiguous sounds, such that fixations on

competitor pictures would increase more strongly across trials in the wide-variance condition than in the narrow-variance condition. However, replicating the pattern reported by Nixon et al. (2016), models including trial number revealed no such interaction here (Experiment 1.1: same-variance, $BF = 0.07$; mixed-variance, $BF = 0.16$; Experiment 1.2: same-variance, $BF = 0.11$; mixed-variance, $BF = 0.08$). These Bayes factors provide moderate-to-strong evidence in favor of the null hypothesis, indicating that competitor fixation did not systematically increase across trials as a function of distributional variance. Taken together, these findings suggest that competitor fixation primarily reflects momentary processing uncertainty, which appears to be more closely related to inconsistency than to gradiency.

Note that the absence of cumulative trial-wise change does not preclude interpreting competitor fixation in terms of gradiency, as gradiency effects may reflect rapid implicit learning that occurs even without observable trial-by-trial change. Combined with the findings of Kapnoula and McMurray (2021), which demonstrate a relationship between lexical gradiency measured by fixation and VAS gradiency, we tentatively suggest that the competitor-fixation measure may provide a more accurate index of gradiency compared to the categorization slope. Focusing on the competitor-fixation findings, we tentatively conclude that gradiency is malleable in bilingual listeners under both same-variance and mixed-variance exposure conditions. However, this effect was observed only when the two languages exhibited substantial VOT overlap (i.e., in the Spanish–Basque case). This pattern may be explained in two ways. First, Spanish–Basque listeners may be able to construct and maintain distinct distributional models for two languages produced by the same talker, allowing them to respond to the wide–narrow manipulation in a manner similar to monolingual listeners. In contrast, Spanish–English bilinguals are routinely exposed to broader VOT distributions due to cross-language differences, which may attenuate the impact of the experimental manipulation. Alternatively, distributional learning may become more

challenging when the VOT distributions of two languages differ markedly, because listeners cannot readily rely on L1-based priors to interpret L2 input. This account aligns with the hierarchical inference framework, in which learning is understood as probabilistic inference under uncertainty: learners integrate incoming linguistic evidence with prior knowledge and update their beliefs about both the structure of the target language and broader phonetic regularities (Pajak et al., 2016). In both accounts, long-term experience with different VOT systems modulates sensitivity to short-term distributional manipulations.

Related to the divergence between measures, although McMurray et al. (2008) did not employ a distributional learning manipulation, they demonstrated that, all else being equal, expanding a 2AFC task to a 4AFC task by adding two fillers resulted in shallower categorization slopes, whereas competitor fixation showed similar gradient patterns across VOT in both task formats. The emergence of shallower slopes following the addition of response options is difficult to reconcile with accounts based solely on either gradiency or consistency. Instead, this pattern suggests that categorization slope is sensitive to task structure and may not provide a pure index of either construct. Together, these findings further indicate that categorization slope and competitor fixation may index partially distinct aspects of speech perception.

Taken together, our findings suggest that exposure to greater acoustic-phonetic variability may enhance listeners' sensitivity to fine-grained distinctions within phoneme categories. Based on this finding, exposure-based learning may be more effective than explicit methods, given that previous studies employing training and instruction have shown that shifting from categorical to gradient perception is quite difficult (e.g., Giovannone et al., 2024; Kapnoula, 2016; Nguyen et al., 2024). Further research is needed to better understand the malleability of gradiency using more stringent methods of exposure. This is an important

issue, as speech perception gradiency has been shown to confer multiple advantages for downstream processing in lexical access, speech-in-noise perception, and language learning (e.g., Bidelman et al., 2025; Kapnoula et al., 2017, 2021; D. Kim et al., 2020; Kong & Edwards, 2016; Kong & Kang, 2023; Lee & Park, 2024; Ou et al., 2021; Rizzi & Bidelman, 2024).

Understanding the effects of acoustic variability in a bilingual context also has practical benefits. Previous studies have shown that increasing variability along an acoustic dimension can lead listeners to rely on that dimension less during subsequent category learning (Holt & Lotto, 2006; Liu & Holt, 2015). Moreover, language training that presents bimodal distributions (e.g., Chládková et al., 2022; Götz et al., 2025; Gulian et al., 2007; Hayes-Harb, 2007; Maye & Gerken, 2000, 2001) or manipulates the variability of acoustic cues (e.g., Escudero et al., 2011; Kondaurova & Francis, 2010; Lim & Holt, 2011; Schreiber et al., 2013; Sumner, 2011) has been found to enhance contrast learning. Together, this body of work highlights the importance of acoustic variability in shaping phonetic learning.

The present study demonstrates that gradiency is malleable in a bilingual context under both same-variance and mixed-variance conditions. We further tentatively suggest that gradiency is more readily malleable when the two languages exhibit greater acoustic-phonetic overlap. At the same time, our results underscore that distributional learning is not exclusively linked to gradiency, but may also relate to consistency. Importantly, this study is the first to demonstrate divergent patterns between categorization slope and competitor-fixation measures within the same distributional learning paradigm. This divergence suggests that different measures may tap into distinct underlying processes, with categorization slope potentially indexing consistency and competitor fixation more closely reflecting gradiency. Further research is needed to clarify how these measures relate to gradiency and consistency, and to distributional learning more generally.

3 Functions of Consistency and Gradiency: Initial Lexical Activation and Speech Perception Flexibility

Parts of this chapter are based on:

- Wong, B. W. L., Samuel, A. G., & Kapnoula, E. C. (in press). The role of speech perception gradiency in L1 versus L2 spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance.* Preprint: https://doi.org/10.31234/osf.io/t5zbc_v5
- Wong, B. W. L., Samuel, A. G., & Kapnoula, E. C. (under review). *Speech perception consistency facilitates initial lexical activation, but not speech perception flexibility.* Preprint: https://doi.org/10.31234/osf.io/gxmnr_v1

3.1 Introduction

This chapter investigates how consistency and gradiency shape initial lexical activation and speech perception flexibility in L1 Spanish and L2 English. As discussed in previous chapters, both consistency and gradiency confer multiple advantages in speech processing and language learning (e.g., Bidelman et al., 2025, 2026; Carraturo et al., 2025; Kapnoula et al., 2017, 2021; H. Kim et al., 2025a–d, 2025f; Myers et al., 2024b; Rizzi & Bidelman, 2025). However, the vast majority of these findings are based on L1 English speakers, leaving open the question of whether similar benefits generalize to other L1s such as Spanish, or extend to L2 processing. Addressing this gap is essential for understanding whether the mechanisms underlying gradiency and consistency are language-general or language-specific, and how they contribute to flexible speech perception across languages.

Converging empirical evidence suggests that gradiency benefits multiple aspects of speech perception. For example, it facilitates the integration of multiple acoustic cues (D. Kim et al., 2020; Kapnoula et al., 2017, 2021; Kong & Edwards, 2016; Kong & Kang, 2023; Lee & Park, 2024; Ou et al., 2021; cf. Kong, 2019). A gradient representation helps listeners remain sensitive to subtle differences within each cue and/or prevents premature commitment to a single cue, thereby allowing more effective integration across cues. Another advantage of gradiency lies in improving speech-in-noise perception (Bidelman et al., 2025, 2026; Rizzi & Bidelman, 2024; cf. Kapnoula et al., 2017, 2021). A gradient perceptual system enables listeners to access more detailed acoustic information in the signal, allowing them to "hedge their bets" when processing speech under conditions of noise or ambiguity. Gradiency has also been linked to L2 learning (Kapnoula & Samuel, 2024). Specifically, evidence suggests that gradiency in a listener's L1 (Spanish) predicts their proficiency in L2 (English), as measured by a vocabulary task. This relationship implies that gradiency may play a crucial

role in L2 acquisition by supporting the establishment of new phonemic categories and maintaining sensitivity to within-category acoustic differences.

In general, higher sensitivity to subphonemic differences may facilitate speech perception through at least three mechanisms. First, more gradient listeners may be better able to utilize subphonemic information to anticipate upcoming input, for instance through cues from coarticulation or assimilation (Gow, 2001; Salverda et al., 2014). Second, maintaining gradient information enables more flexible integration of acoustic cues, as the value of one cue can be interpreted in light of another (Repp, 1982; Summerfield & Haggard, 1977; Winn et al., 2013). Finally, gradient activation of representations across processing levels, such as phonemes and words, allows the system to partially commit to multiple interpretations simultaneously, which is particularly advantageous when encountering highly ambiguous input (Brown-Schmidt & Toscano, 2017; Clayards et al., 2008). In this chapter, we focus on this last potential advantage: speech perception flexibility.

Speech perception flexibility has been extensively examined in listeners' L1, particularly at the phonetic level. For example, ample evidence shows that listeners can adjust their perceptual categories based on lexical information. In the seminal study of lexically guided perceptual recalibration, Dutch listeners were exposed to an acoustically ambiguous sound between /s/ and /f/ embedded in words that could only end with one of those phonemes. Those who heard the ambiguous sound in /s/-final words were subsequently more likely to categorize similar ambiguous sounds as /s/, whereas those who heard it in /f/-final words tended to categorize them as /f/ (Norris et al., 2003). However, much less is known about how this flexibility operates in a bilingual's L2.

The extent to which listeners can adjust their categories in their L2 appears to depend on the difficulty of the specific contrast (e.g., Drozdova et al., 2016; Llompart, 2024; Llompart & Reinisch, 2019; Reinisch et al., 2013; Sjerps & McQueen, 2010). For relatively

easy contrasts, L2 listeners often shift their phonetic boundaries in a manner similar to L1 listeners. For instance, Drozdova et al. (2016) found that Dutch listeners were able to retune L2 English phonetic boundaries after exposure to an ambiguous British English /l/–/ɹ/ sound, with stronger recalibration effects for /ɹ/ than for /l/, possibly due to the greater phonetic variability of /ɹ/ in both Dutch and English. In contrast, for more difficult contrasts, L2 listeners tend not to show the same pattern as in their L1. Llompart and Reinisch (2019) reported that German learners of English who were more successful at accepting words containing /ɛ/ and rejecting nonwords in which /ɛ/ was replaced by the novel L2 vowel [æ] (e.g., *l[æ]mon*) exhibited smaller shifts in their /ɛ/–/æ/ boundary in a distributional learning task. Consistent with this, Llompart (2024) showed that German listeners display reliable perceptual learning effects for an easy L2 contrast that also exists in their L1 (/i/–/ɪ/), but not for a more difficult contrast involving a novel L2 vowel (/ɛ/–/æ/).

While most previous work has focused on flexibility in phonetic categorization, much less is known about flexibility in lexical reinterpretation—the ability to revise an initial word recognition decision when later input disconfirms it. The present study addresses this aspect of speech perception flexibility by targeting a moderately challenging and natural L2 situation: the English voicing distinction between /b/ and /p/ for L1 Spanish listeners. As outlined in Chapter 1, Spanish and English differ markedly in the typical VOTs for labial stops: in Spanish, /b/ ≈ −80 ms and /p/ ≈ 16 ms (Souganidis et al., 2024), whereas in English, /b/ ≈ 10 ms and /p/ ≈ 65 ms (Allen & Miller, 1999; see Figure *1.3*). In Spanish, /b/ is prevoiced, resulting in a negative VOT, while Spanish /p/ aligns more closely with English /b/. According to the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021), L2 sounds that are phonetically close to an existing L1 category are particularly challenging to acquire. Because English /b/ is closer in VOT to Spanish /p/, Spanish listeners should find English /b/ more difficult to learn than English /p/. This contrast is indeed well known to pose

difficulties for Spanish–English bilinguals (Williams, 1977a), making it an ideal test case for examining how gradiency supports word recognition in contexts where L1 and L2 processing are expected to diverge.

To measure speech perception flexibility in lexical reinterpretation, one can examine how listeners cope with misleading or ambiguous auditory cues. For example, McMurray et al. (2009) investigated this process using lexical garden-paths, items that initially sound like one word but ultimately turn out to be another. They tracked lexical activation in real time for such stimuli using the visual world paradigm (VWP; Tanenhaus et al., 1995). This approach allowed the researchers to examine how listeners revised interpretations of misleading information as the speech signal unfolded. Participants heard auditory stimuli such as *þeachball*, in which the initial phoneme varied along a continuum between /b/ and /p/ (with [þ] representing a token from the continuum). As expected, when hearing *peachball*, listeners were initially prone to look at the picture of *peachpit* (the "competitor"), reflecting a *lexical garden-path*. However, as the subsequent input was consistent with *beachball* (the "target"), listeners eventually recovered from the garden-path and recognized *beachball*. The rate (likelihood) and latency (speed) of recovery from lexical garden-paths capture speech perception flexibility. Crucially, in this study, both the garden-path rate and recovery latency were linearly related to the acoustic distance between the stimulus and the target: more competitor-like initial sounds led to more frequent garden-paths and slower recoveries. These findings support the view that listeners maintain gradient representations of speech, which enables them to recover efficiently from initial interpretations caused by ambiguous or misleading input.

While the rate of lexical garden-path responses could indicate the degree of initial lexical activation, it captures only a subset of trials in which participants looked at the competitor. Rather than relying on garden-path rate, we included all trials to index initial

lexical activation. Specifically, initial lexical activation was operationalized as early fixations on /b/- and /p/-onset referents (e.g., *beachball* and *peachpit*) as a function of VOT. For this measure, we included only fixations occurring before the point of disambiguation (POD), defined as the point in the auditory signal at which the input uniquely specifies the target word (e.g., the transition from *beach* to *ball* in *beachball*). Increased fixations on the /b/-onset referents when VOT was low and increased fixations on the /p/-onset referents when VOT was high were taken to indicate stronger initial lexical activation.

Combining the VWP with VAS measures, Kapnoula et al. (2021) examined the link between gradiency and flexibility in resolving lexical garden-paths. They found that more gradient listeners were better able to recover from misleading or ambiguous input. Kapnoula et al. (2021) argued that this advantage arises because gradient listeners maintain multiple interpretations in parallel, avoiding premature commitment to a possibly incorrect lexical choice. These results suggest that gradient speech perception provides listeners with greater flexibility to revise their interpretations when faced with ambiguous or conflicting information.

While Kapnoula et al. (2021) provided evidence that gradiency facilitates speech perception flexibility in L1, much less is known about whether gradiency plays a comparable role in L2 speech perception flexibility. Addressing this question not only can extend our understanding of how gradiency shapes speech perception, but can also provide important theoretical implications grounded in cross-linguistic differences between L1 and L2 processing. For instance, the relative reliance on bottom-up versus top-down information may differ across a bilingual's two languages, potentially modulating the contribution of gradiency, which primarily reflects bottom-up processing.

While listeners draw on both bottom-up (e.g., VOT, vowel length, speech rate) and top-down (e.g., semantic, lexical, gender marking) information during speech perception,

individuals differ in the extent to which they rely on one source over the other (e.g., Crinnion et al., 2024; Giovannone & Theodore, 2023; Kaufeld et al., 2020). Listeners who rely more on high-level information tend to depend less on acoustic-phonetic cues, and vice versa. This differential reliance is influenced by several factors. For instance, reliance on lexical information increases under cognitive load (Derawi et al., 2022; Mattys et al., 2009, 2010; Mattys & Wiget, 2011; but see Mattys et al., 2014), anxiety (Mattys et al., 2013), and with aging (Mattys & Scharenborg, 2014). Similarly, cochlear implant users and individuals perceiving degraded speech (Gianakas & Winn, 2019), as well as children with developmental dyslexia or specific language impairments, show stronger lexical reliance (Derawi et al., 2022; Reed, 1989; Schwartz et al., 2013; Tufo & Myers, 2014). Moreover, lower-proficiency listeners tend to rely less on bottom-up acoustic-phonetic cues and more on top-down lexical information (Giovannone & Theodore, 2021a, 2021b). Although these studies did not directly compare L1 and L2 processing, they collectively suggest that when speech perception becomes more difficult, as is often the case in L2 listening, listeners exhibit greater reliance on top-down (lexical) information.

However, when directly comparing bottom-up and top-down reliance in L1 and L2 speech perception, findings do not always align with this interpretation. Several studies suggest that L2 listeners may, in some cases, rely less on top-down lexical cues, likely due to having weaker lexical–semantic representations (Ishida & Arai, 2016; Mattys et al., 2010; Samuel & Frost, 2015). For instance, Samuel and Frost (2015) found that lower-proficiency L2 listeners relied less on lexical information than both L1 and higher-proficiency L2 listeners. They proposed that as L2 proficiency increases, listeners develop more robust lexical representations that can subsequently be leveraged to facilitate speech processing.

Overall, the evidence suggests that reliance on top-down information is generally higher in L2 than in L1, but only when such information is sufficiently well established,

explaining the lower lexical reliance observed in low-proficiency L2 listeners. If L2 speech perception indeed depends more heavily on top-down information than L1 perception, given that gradiency primarily reflects bottom-up processing, the influence of gradiency should therefore be weaker in L2 than in L1.

Our first research objective was to examine whether gradiency predicts initial lexical activation and speech perception flexibility in both L1 and L2. To this end, the same group of participants completed two tasks: a VAS task assessing speech perception gradiency and a VWP task assessing initial lexical activation and speech perception flexibility in Spanish (L1) and English (L2). As detailed in Sections 3.3.1 and 3.3.2, these two tasks provide insight into cross-linguistic differences in bottom-up and top-down reliance, and how such differences may modulate the role of gradiency in speech perception and lexical processing.

Following Kapnoula et al. (2021), we predicted that in L1, greater within-category sensitivity would facilitate lexical recovery, but would not be directly related to initial lexical activation. Accordingly, we expected the benefit of gradiency for speech perception flexibility to be replicated in L1 Spanish. As discussed above, the role of gradiency may be modulated by cross-linguistic differences in bottom-up and top-down processing. Thus, if L2 speech processing relies more heavily on top-down lexical information and less on bottom-up acoustic-phonetic cues, which may be due to the perceptual difficulty of the /b/–/p/ contrast, the impact of gradiency on L2 English should be attenuated.

In addition to gradiency, another important construct that may shape speech perception is consistency. Although the advantages of consistency have been demonstrated across multiple domains (e.g., Carraturo et al., 2025; Fuhrmeister et al., 2023; Honda et al., 2024a), it remains largely unknown whether and how consistency contributes to initial lexical activation and speech perception flexibility. Therefore, another primary research objective was to examine the relationship between consistency and both initial lexical activation and

speech perception flexibility. We predicted that listeners who exhibit low speech perception consistency would demonstrate lower initial lexical activation of input-compatible candidate words. This prediction is grounded in previous findings showing that listeners tend to delay lexical activation when the input is degraded or noisy (e.g., cochlear implant users and normal-hearing listeners presented with noise-vocoded speech; Farris-Trimble et al., 2014; McMurray et al., 2017; Smith & McMurray, 2022). Listeners with lower consistency may experience greater variability or lower stability in their cue-to-category mappings, leading to increased uncertainty during speech processing. This uncertainty may induce a "wait-and-see" strategy in the early stages of spoken word recognition, whereby initial lexical activation is slowed to avoid premature commitment to a lexical candidate until additional information becomes available. Based on this rationale, the predicted relationship between consistency and initial lexical activation should hold in both L1 and L2, as the proposed mechanism, greater uncertainty arising from lower stability in cue-to-category mapping, is expected to influence lexical processing regardless of language. Moreover, we did not expect consistency to play a major role in speech perception flexibility in either L1 Spanish or L2 English, as flexibility reflects the ability to adapt and recover from misleading information rather than the stability of phonetic representations per se.

Our third research objective was to examine the relationship between L1 Spanish and L2 English gradiency, as well as between L1 Spanish and L2 English consistency, in order to better understand the characteristics of these two constructs. To our knowledge, only a few studies have investigated gradiency across both L1 and L2 speech contrasts (Desmeules-Trudel & Zamuner, 2023; Kong, 2019; Kong & Kang, 2023). Specifically, Kong (2019) tested English learners of L2 Korean on gradiency in L1 English (/d–t/) and L2 Korean (lax /t/–aspirated /th/), whereas Kong and Kang (2023) examined Korean learners of L2 English using the same stimuli. In the first case, a positive but nonsignificant correlation was

observed between L1 and L2 gradiency; in the second, L1 gradiency explained a significant portion of L2 gradiency (R^2 = .309). In addition, Desmeules-Trudel and Zamuner (2023) showed that both L1 French speakers and L1–L2 English–French bilinguals exhibit gradient sensitivity to vowel nasalization in French, although they did not explicitly test correlations between L1 and L2 gradiency. While consistency has been shown to be a general trait across contrasts and modalities (e.g., visual and auditory; H. Kim et al., 2025d, 2025e), to our knowledge, no study has yet examined consistency across languages. Based on the view that consistency reflects a stable individual trait, we predicted a positive correlation between L1 and L2 consistency, but a weaker correlation between L1 and L2 gradiency.

To address these three research objectives, we tested gradiency, consistency, initial lexical activation, and speech perception flexibility in L1 Spanish speakers, both in their L1 (Spanish) and their L2 (English).[4] We focused on the /b/–/p/ contrast in both languages. Participants also completed a picture-naming task to assess English proficiency. In addition, we included English exposure, working memory, inhibitory control, and musical training as covariates, given their previously reported links to language processing and gradiency (e.g., Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; Kapnoula et al., 2017; D. Kim et al., 2020; Kong, 2019; McMurray et al., 2018; Smayda et al., 2015).

3.2 Method

3.2.1 Participants

We recruited 70 L1 Spanish speakers (53 females; 17 males) residing in San Sebastian, Spain for this study. This sample size was informed by a power analysis (see

[4] English is referred to as L2 here, while acknowledging that for many participants, English may function as their second, third, or even subsequent language. Recall that most participants residing in the Gipuzkoan province of the Basque Country also speak Basque. However, since the VOT of /b-p/ in Basque closely resembles that of Spanish (Souganidis et al., 2024), and Basque exposure did not introduce additional variance into our models, we believe Basque does not play a significant role in our findings and will therefore not be discussed further.

Appendix 1.2). Participants were between 18 and 40 years old (M = 27.8; SD = 6.8) and reported normal or corrected-to-normal vision and no history of hearing, reading, motor, or speech disorders, as well as no history of head or brain surgery, claustrophobia, or color blindness. All participants acquired Spanish between birth and age six (AoA = 0–6)[5] and identified Spanish as their dominant language. To be included, participants were required to achieve Spanish proficiency scores of at least 92% on the picture-naming task. In addition, their English proficiency scores ranged from 23% to 100%, as we deliberately included participants with a wide range of L2 proficiency levels to capture individual differences. Detailed demographic information is provided in Table A3.1. The study was approved by the BCBL Ethics Review Board (ref. 030423SM) and adhered to the principles of the Declaration of Helsinki. All participants provided written informed consent and received monetary compensation for their participation.

3.2.2 Task Overview

We administered the VAS to extract measures of participants' speech perception gradiency and consistency. To measure initial lexical activation and speech perception flexibility (i.e., recovery rate and recovery latency), we employed a VWP task with eye-tracking, similar to those used by Kapnoula et al. (2021) and McMurray et al. (2009). The VAS and VWP tasks were administered in English (L2) with the same stimuli and design as in Kapnoula et al. (2021); both tasks were also conducted in Spanish (L1) with Spanish stimuli in a separate session. The order of sessions was counterbalanced (see Section 3.2.9 for details). Additionally, participants completed an English picture-naming task (assessing English proficiency), a spatial Stroop task (assessing inhibitory control), a Corsi block-tapping task (assessing working memory), and a questionnaire on language and musical

[5] All participants except two had a Spanish AoA of 0. Given that these two participants reported Spanish as their dominant language and had higher Spanish proficiency than Basque or English proficiency, they were included in the analyses.

background. All tasks were conducted in a sound-attenuated booth. The tasks and variables are summarized in Table *3.1*.

Table 3.1
Summary of Tasks and Variables for Experiment 2

Task	Variable	Duration (min)
Spanish and English VAS	Speech perception gradiency and consistency	8 for each language
Spanish and English VWP	Initial lexical activation and speech perception flexibility (i.e., recovery rate and recovery latency)	52 for each language
English picture naming task	English proficiency	5
Spatial Stroop task	Inhibitory control	5
Corsi Block-tapping task	Working memory	5
Questionnaire	Language and musical background	3

Note. VAS = Visual Analog Scale; VWP = Visual World Paradigm.

3.2.3 VAS

The English VAS stimuli were identical to those used in Kapnoula et al. (2017). Specifically, we used a "buh"–"puh" continuum consisting of natural speech items recorded by a male L1 speaker of American English that varied factorially along (a) seven VOT steps (1–45 ms) and (b) five F_0 steps (90–125 Hz). Similarly, the Spanish stimuli formed a natural speech "ba"–"pa" continuum with seven VOT steps from −35 to +10 ms (−35, −28, −20, −12, −5, +3, +10 ms) and five F_0 steps from 179 to 193 Hz (179, 183, 186, 189, 193 Hz). While the English stimuli were generated using progressive cross-splicing (see Kapnoula et al., 2017, for details), the Spanish stimuli were manipulated using the VOT-synthesis script by Winn (2020; version 33). The original natural recordings used as endpoints had VOTs of −58 ms and +25 ms.

As described in Chapter 1, in the VAS, participants viewed a horizontal line labeled /ba/ on the left and /pa/ on the right (see Figure *1.1*). On each trial, they heard one of the

stimuli described above and indicated where they perceived the sound to fall along the continuum by clicking on the line. After their first click, a rectangular bar appeared at the selected point, and participants could modify their response or press the space bar to confirm it. There was no time limit for making their response. The task consisted of three blocks of 35 trials, yielding a total of 105 trials (5 F_0 × 7 VOT × 3 repetitions), with stimuli presented in random order. In addition, five practice trials were presented at the beginning to ensure that participants understood the task. Participants were informed about the language of the task (English or Spanish) before it began. This task took around eight minutes for each language.

3.2.4 VWP

In each language, the VWP stimuli consisted of word pairs that differed in initial voicing (/b/ vs. /p/). These pairs shared similar onset phonemes, differing only in the voicing of the first consonant, contained two to five shared medial phonemes, and had distinct offsets (e.g., *beachball* [bitʃbɔl] vs. *peachpit* [pitʃpɪt]). Each pair underwent VOT manipulation to create a seven-step continuum from word to nonword forms (e.g., *beachball* [bitʃbɔl] to *peachball* [pitʃbɔl]; *peachpit* [pitʃpɪt] to *beachpit* [bitʃpɪt]). Instances of initial VOT ambiguity were disambiguated by phonetic information at the word offset (e.g., *–ball* or *–pit*), progressively increasing the likelihood that participants would initially activate the competitor word and enabling us to evaluate whether and how quickly they recovered upon hearing the disambiguating offset (reflecting speech perception flexibility). The English words and pictures, including five pairs of critical stimuli beginning with /b/ and /p/ and five filler pairs beginning with /l/ and /r/, were identical to those developed by Kapnoula et al. (2021) (see Table *3.2* for details of all English stimuli). Since the design and materials of the English task are described in Kapnoula et al. (2021), the following section focuses on the Spanish VWP.

Table 3.2
Stimuli Used in the English VWP with International Phonetic Alphabet (IPA), as in Kapnoula et al. (2021)

	Critical Stimuli				Filler Stimuli			
Set	Voiced word		Voiceless word		Lateral approximant word		Rhotic approximant word	
	Spelling	IPA	Spelling	IPA	Spelling	IPA	Spelling	IPA
1	bumpercar	bʌmpər**kɑr**	pumpernickel	pʌmpər**nɪkəl**	lemonade	lɛməˈneɪd	restaurant	rɛstᵊrɒnt
2	barricade	bærək**eɪd**	parakeet	pærək**it**	limousine	lɪməˈziːn	raspberry	rɑːzbəri
3	blanket	blæŋk**ɪt**	plankton	plæŋk**tən**	lantern	læntən	rabbit	ræbɪt
4	beachball	bitʃ**bɔl**	peach-pit	pitʃ**pɪt**	lobster	lɒbstə	ribbon	rɪbᵊn
5	billboard	bɪlb**ɔrd**	pillbox	pɪlb**ɒks**	lemon	lɛmən	rhino	raɪnəʊ

Note. Underlined portions mark phonemic overlap between the words in a pair for critical stimuli; bolded portions mark offsets.

Table 3.3
Stimuli Used in the Spanish VWP with IPA and Meanings

	Critical Stimuli						Filler Stimuli					
Set	Voiced word			Voiceless word			Lateral approximant word			Rhotic trill word		
	Spelling	IPA	Meaning	Spelling	IPA	Meaning	Spelling	IPA	Meaning	Spelling	IPA	Meaning
1	balanza	bala**nθa**	scale	palacio	pala**θjo**	palace	lavabo	laβaβo	sink	regalo	reɣalo	gift
2	bañar	baɲa**r**	to bathe	pañal	paɲa**l**	diaper	ladrón	laðɾon	thief	reptil	reptil	reptile
3	baraja	baɾax**a**	deck of cards	paraguas	paɾaɣ**was**	umbrella	ladrillo	laðɾiʎo	brick	ropero	ropeɾo	wardrobe
4	vaquero	bake**ro**	cowboy	paquete	pake**te**	parcel	linterna	linteɾna	flashlight /torch	rosario	rosaɾjo	rosary
5	vestido	best**iðo**	dress	pestañas	pest**aɲas**	eyelash	lechuga	letʃuɣa	lettuce	retrato	retɾato	portrait

Note. Underlined portions mark phonemic overlap between the words in a pair for critical stimuli; bolded portions mark offsets.

For the Spanish stimuli, we developed a set of 20 Spanish words, including critical stimuli consisting of five /b/-onset and five /p/-onset words, as well as five filler word pairs with /l/- and /r/-onsets (see Table *3.3* for details of all Spanish stimuli). The characteristics of these stimuli, along with their corresponding pictures, mirrored those of the English set. Each word contained five to eight phonemes and two or three syllables, with three phonemes overlapping within each /b/–/p/ word pair.

The recording procedure and stimulus processing were similar to those described in Section 2.2.3. All items, including the real words (e.g., *balanza*, *palacio*) and their corresponding nonwords (e.g., *palanza*, *balacio*), were recorded by a female speaker of Peninsular Spanish from San Sebastian. Each word was produced within the carrier phrase *Es mi* ___ ("It's my___") to maintain consistent prosody, and four repetitions of each sentence were recorded. The recordings were denoised in Audacity (Version 2.3.3; Audacity Team, 2019). Target words were then excised from the carrier sentence, and the most natural and clearly articulated token of each word was selected. All sound files were normalized to 70 dB using Praat (Version 6.3.03; Boersma & Weenink, 2023). Each recording was then divided into two segments: the onset (e.g., "bala-" from *balanza*) and the offset (e.g., "–nza"). The stimuli were segmented at the zero-crossing point nearest to the point of disambiguation (POD), that is, the moment in the auditory signal when the input uniquely identifies the target word (e.g., the boundary between *beach* and *ball* in *beachball*). The average POD occurred at 433 ms.

Given that the onset portions could contain coarticulatory cues predicting the offset (e.g., the "bala–" from *balanza* may predict "–nza" more than "–cio"), each of the two voiced onsets in a pair (e.g., "bala$_{nza}$" and "bala$_{cio}$") was spliced onto each of the two offsets to counterbalance the coarticulatory cues in the onsets (e.g., "$_{bala}$nza" and "$_{bala}$cio"). Consequently, half of the stimuli were generated by combining parts from the same item

(e.g., "bala$_{nza}$" and "$_{bala}$nza"; *matching splice*), while the other half were formed from different items (e.g., "bala$_{nza}$" and "$_{pala}$nza"; *mismatching splice*). Further details are provided in Figure S4 in the Supplemental Materials of Kapnoula et al. (2021).

The VOT continua were constructed by pairing items that differed solely in the voicing of the onset consonant (e.g., *balanza* and *palanza*). These pairs served as the endpoints for generating seven-step VOT continua (Spanish: −35 to +15 ms; English: 0 to +48 ms). The construction procedure followed the method described in Section 2.2.3. This process yielded 140 auditory items per language (5 pairs × 2 splice conditions × 2 offsets × 7 VOT steps). The F_0 was standardized at 181 Hz for all Spanish stimuli. Finally, 100 ms of silence was added to both the onset and offset of each token. Each item was presented three times, resulting in 420 experimental trials.

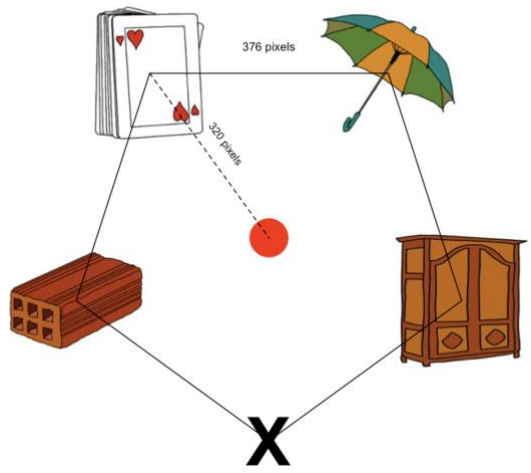
For filler stimuli, only the correct (e.g., *lavabo* "sink" and *regalo* "gift") and misarticulated (e.g., *ravabo* and *legalo*) versions were recorded and used. An equivalent number of filler trials (i.e., 420 trials) were interspersed among the critical trials to introduce variety and obscure the task purpose, yielding a total of 840 trials. All trials were presented in a randomized order.

Each experimental pair was paired with a filler pair to create a four-item set (e.g., *balanza*, *palacio*, *lavabo*, and *regalo* formed one set in the Spanish VWP), with the constraint that all items within a VWP set were semantically unrelated and shared the same stress pattern. Items belonging to a given VWP set always appeared together. Five VWP sets were created for each language.

A picture was selected for each of the four words in each of the five VWP sets. The procedure for selecting and fine-tuning the pictures followed the steps described in Section 2.2.3.

Experiment Builder (version 2022.2.5; SR Research Ltd.) was used to program and run the experiment, with identical settings across the two languages. Visual stimuli were presented on a 24-inch monitor operating at a resolution of 1024 × 768 pixels. Each display contained five visual elements, including four pictures and an X, arranged in a pentagonal configuration (see Figure *3.1*). This configuration ensured equal distances between the center of each picture and the screen center (320 pixels), as well as between adjacent pictures (376 pixels apart). Each picture measured 240 × 240 pixels, and the X measured 66 × 80 pixels. The positions of the four pictures were randomized across trials, whereas the X was always positioned at the bottom.

Figure 3.1
VWP Display



Note. Lines and pixel values are for illustration and were not shown during the experiment. Images shown here were from the Spanish VWP; the same configuration was used in the English VWP.

The procedure was similar to that of the VWP described in Section 2.2.4. Participants first completed a familiarization phase, during which each picture and its corresponding label were presented individually. Participants were then fitted with the eye tracker, completed calibration, and received task instructions. At the beginning of each trial in the main experiment, a red circle appeared at the center of the screen together with four pictures and an

X arranged in a pentagonal configuration. After 500 ms, the circle turned blue, and participants clicked on it to hear a word. This brief pause allowed participants to preview the pictures before the auditory stimulus onset, thereby minimizing eye movements driven by visual search (Apfelbaum et al., 2021). Participants then clicked on the corresponding picture, or on the X if they believed none of the pictures matched the word. The trial ended once a response was made, and the next trial began immediately.

Eye-tracking recording and analysis were the same as those described in Section 2.2.5, except that an older version of EyelinkAnalysis (Version 3.5.1; McMurray, 2025) was used for eye-tracking data preprocessing. Picture boundaries were extended by 58 pixels to accommodate potential noise or minor head drift in the eye-tracking record. This extension did not produce any overlap, as a vertical separation of 118 pixels and a horizontal separation of 136 pixels were maintained between pictures.

3.2.5 English Picture Naming Task

Expressive vocabulary was used as an index of English proficiency. We employed an English picture-naming task (de Bruin et al., 2017) featuring 65 pictures corresponding to noncognate words, drawn from the MultiPic database (Duñabeitia et al., 2018). Participants viewed each picture and typed a single English word to identify it. The pictures represented diverse semantic categories such as animals (24 items) and body parts (eight items). Each response was carefully reviewed: number-agreement errors (e.g., “eyes” instead of “eye”) and typos caused by pressing an adjacent key (e.g., “keg” instead of “leg”) were counted as correct. Participants received a score ranging from 0 to 65 based on the number of correctly named pictures. The task was implemented in jsPsych (Version 7.3.1; de Leeuw et al., 2023) and administered via the Chrome browser.

3.2.6 Spatial Stroop Task

To assess inhibitory control independently of language, we implemented a spatial variant of the Stroop task (Jensen & Rohwer, 1966; Stroop, 1935; Wühr, 2007; also used in Kapnoula & McMurray, 2021). In this task, participants were presented with an arrow displayed either on the left or right side of the screen, pointing left or right. Their response was based on the direction of the arrow, with the irrelevant cue being the arrow's on-screen location. Participants were expected to respond faster and more accurately when the arrow's direction was congruent with its location. To enhance the congruency effect, we used 64 congruent and 32 incongruent trials (Logan & Zbrodoff, 1979), presented in a pseudorandomized sequence. The magnitude of the congruency effect, calculated as the difference in RT between congruent and incongruent trials, excluding incorrect responses, served as an index of participants' ability to suppress irrelevant information, namely the arrow's location. Each trial began with a fixation point presented at the screen center for 200 ms. The fixation point then disappeared, and an arrow appeared on either the left or right side of the screen. Participants pressed one of two keys (left or right) to indicate the arrow's direction. Following the response, a 1,000-ms blank screen preceded the next trial. The task was implemented in Experiment Builder (version 2022.2.5; SR Research Ltd.) and lasted approximately five minutes.

3.2.7 Corsi Block-Tapping Task

To assess visual working memory, we employed a Corsi Block-tapping task (Corsi, 1972; Milner, 1971), implemented using OpenSesame v.3.3.14 (Mathôt et al., 2012). In each trial, nine purple squares were irregularly distributed on a black background. The task began with two practice trials in which two squares sequentially turned green for one second each in a random order. Participants were instructed to memorize the sequence and then click on the squares in the reverse order. Successful completion of both practice trials was required to

proceed to the main experiment. In the main task, the procedure was identical to that of the practice phase, except that sequence length varied from three to nine items. Each sequence length comprised two trials, and participants had to complete at least one of the two trials correctly to advance to the next length. If both trials were answered incorrectly, the experiment terminated. The working memory score corresponded to the maximum sequence length with at least one correct trial. The task lasted approximately five minutes.

3.2.8 Online Questionnaire and Training

To ensure participants' familiarity with the English words used in the English VWP, an online training session was scheduled one to seven days before the in-lab English VWP task. Before the training, participants completed a language exposure questionnaire adapted from Freed et al. (2004), which focused on their English listening experiences. They reported the percentage of time they were exposed to Spanish, English, and Basque across hearing, reading, writing, speaking, and overall. Participants also reported the duration of any stays in English-speaking countries and rated the percentage of time they spent listening to English. In addition, the questionnaire included items about the experience and duration of formal musical training. Finally, a question about hearing problems was included to ensure that this information was up to date at the time of testing (no participant reported any hearing problems). The questionnaire took approximately three minutes to complete.

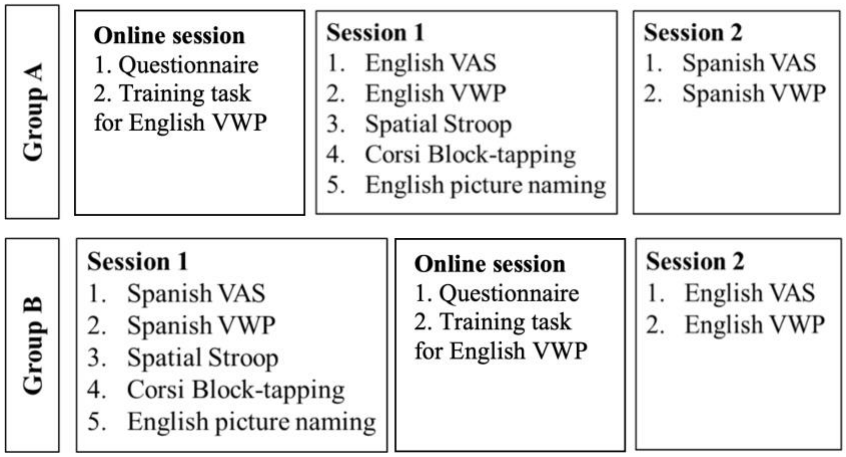
Before the online training began, participants received written instructions recommending the use of headphones and ensuring that their sound was turned on. This was followed by an audio check phase, during which a female voice repeatedly saying "42" (in Spanish) was played to allow participants to adjust the volume. To confirm that they could hear the sound clearly, participants were required to enter the number "42" before proceeding to the training. The first part of the online training was a familiarization phase, in which

participants viewed each image individually, accompanied by the corresponding English word displayed on the screen and simultaneously presented auditorily. They then proceeded to the training phase, where each English word appeared above two pictures: one corresponding to a critical item (e.g., *beachball*, *peachpit*) and the other to a filler (e.g., *lobster*, *ribbon*) from the English VWP. Participants selected one of the two pictures. After each response, a message appeared in the center of the screen for 750 ms indicating whether the response was correct ("¡Correcto!") or incorrect ("Incorrecto."), followed by a 250-ms blank screen. The next trial then began. The training included 20 stimuli (10 critical and 10 filler words). Each stimulus appeared once per block, and across five blocks, participants completed 100 trials presented in a pseudorandomized order. Participants were to be excluded if they failed to answer all questions correctly in at least two of the five blocks. All participants met this criterion, achieving a minimum accuracy of 97 out of 100. Thus, no participants were excluded based on this performance measure.

3.2.9 Order of Tasks

Participants were randomly assigned to either Group A ($N = 35$) or Group B ($N = 35$). The order of tasks for each group is shown in Figure 3.2. All participants attended two laboratory sessions scheduled seven to 14 days apart. The first session lasted approximately 1.5 hours, and the second session lasted about one hour. In addition, as described above, all participants completed an online questionnaire and a training task designed to familiarize them with the English stimuli one to seven days prior to their English session.

Figure 3.2
Order of Tasks for Each Group

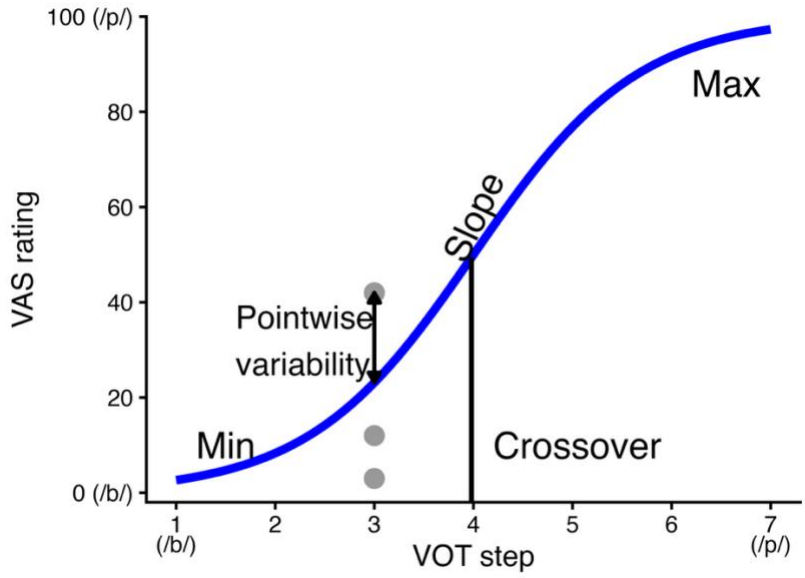


3.2.10 Extracting Speech Perception Gradiency and Consistency

First, click locations along the VAS x-axis were converted from pixel values ranging from 100 to 924 into VAS ratings on a 0–100 scale. To extract gradiency and consistency from participants' responses in the VAS task, curve fitting was conducted largely following the procedure described in Section 2.2.7. However, because both VOT and F_0 were manipulated in this task, gradiency was quantified by fitting a rotated logistic function. Unlike standard logistic regression, this approach provides orthogonal measures of gradiency and secondary cue use (i.e., reliance on F_0). A four-parameter logistic function was fitted to each participant's response curve to estimate the minimum and maximum asymptotes, slope (which is inversely related to gradiency), and crossover point based on their average VAS responses (y) across VOT steps (x_1) and F_0 steps (x_2).

In addition, pointwise variability was calculated as a measure of consistency (see Figure *3.3* for an illustration of the parameters). First, residuals were computed as the difference between the VAS rating on each trial and the predicted value for that stimulus based on the participant's fitted curve. The standard deviation of these residuals was then calculated for each participant. Larger standard deviations indicate lower consistency.

Figure 3.3
Illustration of Minimum and Maximum Asymptotes, Slope, Crossover, and Pointwise Variability Extracted from the VAS Task



Note. The curve represents the predicted (rotated) logistic function obtained from curve fitting, and the dots indicate individual trial responses. Slope and pointwise variability are the inverse measures of gradiency and consistency, respectively.

To ensure that a shallow slope reflects gradiency rather than noisy encoding, consistency was accounted for in the gradiency measure. Specifically, a linear regression was conducted with the log-transformed VAS slope as the dependent variable and consistency as the predictor, and the standardized residuals were extracted. These residuals were used as the gradiency measure in subsequent analyses. Because higher slope and greater pointwise variability correspond to lower gradiency and lower consistency, respectively, both slope and variability measures were multiplied by −1 so that higher values reflect greater gradiency and consistency.

3.2.11 Statistical Analyses

We mainly used Bayesian linear mixed-effects models (LMEMs), as described in Section 2.2.6 to examine our research questions. Based on Bayesian model comparison using predictive accuracy via leave-one-out cross-validation (LOO), inhibitory control, working

memory, English listening exposure, and musical training duration were excluded from the models, and the optimal random-effects structure was determined. Given the number and complexity of the analyses, details of variable transformations and model specifications are reported together with the corresponding results below.

3.3 Results

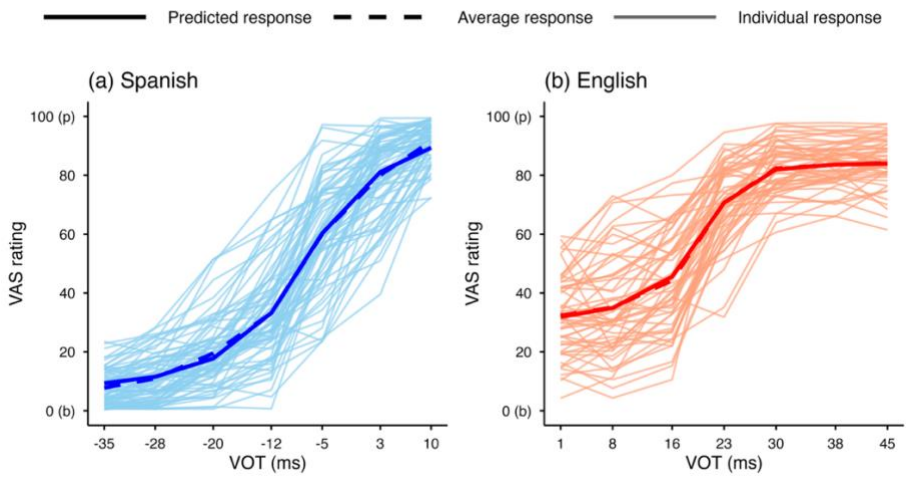
All 70 participants completed all tasks. However, data were excluded from five participants in the Spanish VWP and from four participants in the English VWP because they fixated on the target picture in fewer than 60% of trials within the 300-ms window preceding their click on the target picture. For the English VAS, one participant was excluded for exhibiting a flat response curve (i.e., identical responses regardless of stimulus variation from /b/ to /p/), one for a reversed response curve (i.e., a decrease in the proportion of /p/ responses as stimuli changed from /b/ to /p/), and 12 due to problematic fits ($R^2 < .7$) resulting from a restricted VAS response range. No participants were excluded from the Spanish VAS. The remaining fits were good, with an average R^2 of .940 and .893 for the Spanish and English VAS tasks, respectively. All data from the picture-naming task were retained. To maximize statistical power, all remaining data were included in the analyses. Consequently, the number of participants included in subsequent analyses varied from 52 to 66.

3.3.1 Preliminary Analyses: VAS

Figure *3.4* presents the average VAS data for Spanish and English. As shown, the expected response patterns were observed in the VAS tasks, such that VAS ratings shifted progressively toward /p/ across VOT steps. However, there was substantial inter-individual variability in the degree of gradiency. Some participants selected intermediate points along the response line more frequently, indicating higher sensitivity to within-category variability, whereas others showed a pronounced preference for the endpoints, reflecting more categorical judgments. Participants also differed in their consistency across identical stimuli.

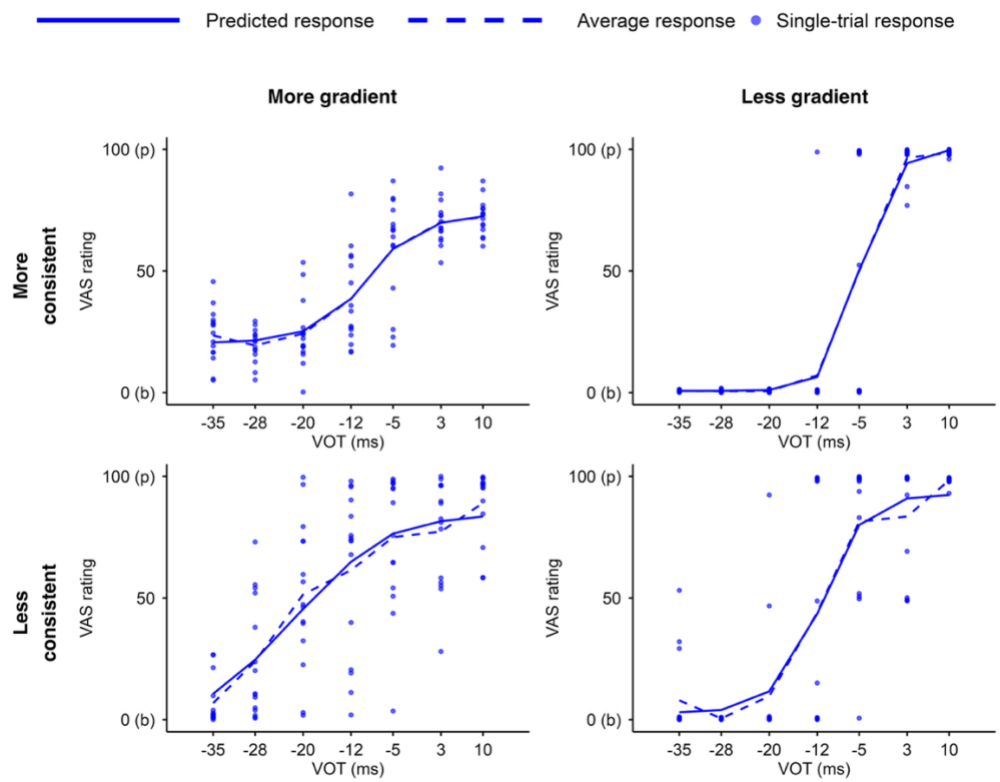
To illustrate these differences, Figure *3.5* (identical to Figure *1.2*) presents representative examples of Spanish VAS data from individual participants exhibiting distinct gradiency and consistency profiles, while Figure *3.6* displays two examples of English VAS data with differing gradiency patterns. Notably, the bottom-left panel of Figure *3.5* demonstrates that a participant can exhibit gradient responses while being inconsistent (i.e., showing noisy encoding), underscoring the importance of considering consistency when quantifying gradiency.

Figure 3.4
VAS Rating across VOT for (a) Spanish and (b) English



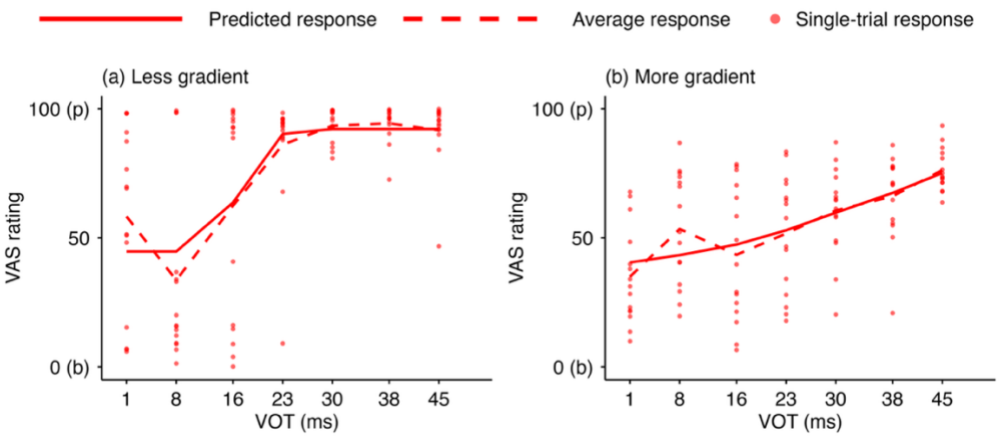
Note. Each light-colored curve represents the average of a participant's response at each VOT value. The solid dark line is the fitting curve based on the rotated logistic function, while the dashed dark line connects the grand average of all participants' responses.

Figure 3.5
Examples of Spanish VAS Data Showing Different Gradiency and Consistency of Four Participants



Note. Each dot represents a participant's response on one trial. The solid line is the fitting curve based on the rotated logistic function, while the dashed line connects the average response at each VOT value for that participant. This figure is the same as Figure *1.2*.

Figure 3.6
Examples of English VAS Data Showing (a) Lower Gradiency and (b) Higher Gradiency in Two Participants



Note. Each dot represents a participant's response on one trial. The solid line is the fitting curve based on the rotated logistic function, while the dashed line connects the average response at each VOT value for that participant.

Figure *3.4* illustrates clear differences between L1 and L2 speech processing. While participants used nearly the entire VAS response range for L1 Spanish (Figure *3.4*a), the same individuals exhibited a markedly narrower response range for L2 English (Figure *3.4*b). This pattern suggests that category representations in L2 English may be less robust than those in L1 Spanish. To test this directly, paired-sample t-tests were conducted on the upper and lower asymptotes derived from the rotated logistic regression fits, as well as on the response range, defined as the difference between the upper and lower asymptotes.

The English lower asymptote ($M = 29.70$, $SD = 12.64$) was higher than the Spanish lower asymptote ($M = 8.01$, $SD = 6.04$), $t(55) = 11.67$, $BF = \infty$, indicating that participants had less robust category representation for English /b/, compared to Spanish /b/. Similarly, the English upper asymptote ($M = 84.62$, $SD = 7.33$) was lower than the Spanish upper asymptote ($M = 94.35$, $SD = 6.58$), $t(55) = -9.56$, $BF = \infty$, suggesting a similarly less robust category representation for English /p/. Based on these results, it is not surprising that the overall response range was also smaller for English ($M = 54.93$, $SD = 14.70$) than for Spanish ($M = 86.34$, $SD = 10.25$), $t(55) = -15.21$, $BF = \infty$. All these differences showed decisive Bayesian evidence. Taken together, these results indicate that listeners exhibited less robust category representations for both English /b/ and /p/ relative to their L1 Spanish.

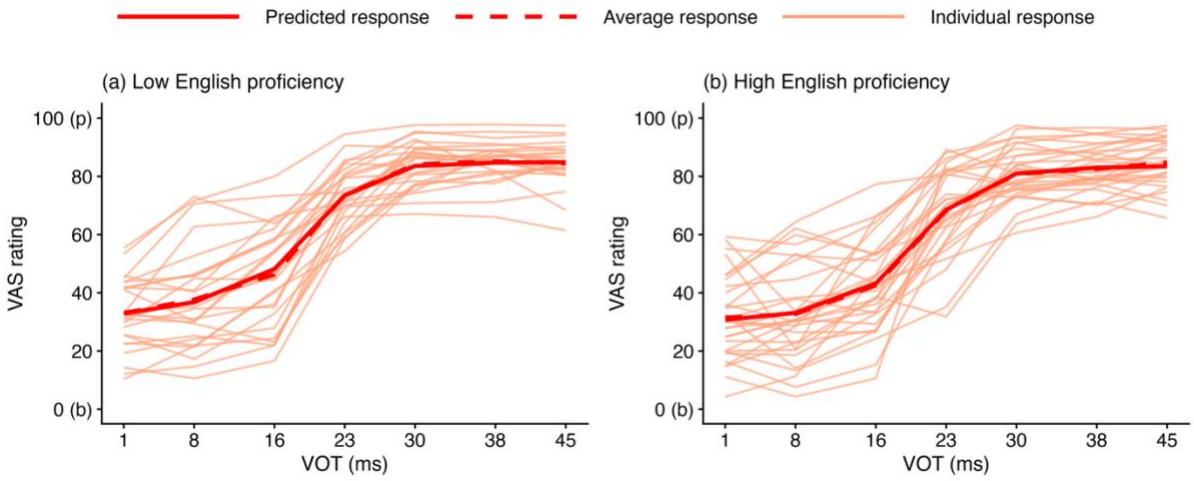
Then, we tested whether listeners exhibited less robust category representations for English /b/ than for English /p/. As was illustrated in Figure *3.4*b, this appeared to be the case: the distance between the lower asymptote and 0 (the minimum possible response; $M = 29.70$, $SD = 12.64$) was greater than the distance between the upper asymptote and 100 (the maximum possible response; $M = 15.38$, $SD = 7.33$). To evaluate this difference statistically, we conducted a paired-samples t-test comparing the two distances. The results revealed decisive evidence that the distance from the lower asymptote to 0 was greater than that from

the upper asymptote to 100, $t(55) = 7.38$, $BF = \infty$. These findings suggest that participants possessed more robust category representations for English /p/ than for English /b/, consistent with the fact that the VOT of Spanish /p/ closely resembles that of English /b/. Importantly, the same English VAS stimuli were used in Kapnoula et al. (2017), whose L1 English participants exhibited a similarly wide response range to that of our L1 Spanish group. This rules out the possibility that the observed effect reflects stimulus characteristics and reinforces the interpretation that it arises from L2 processing limitations.

We also compared gradiency (unresidualized) and consistency between the Spanish and English VAS tasks to validate our English gradiency measure. There was strong evidence that the English gradiency ($M = -1.80$, $SD = 0.40$) was lower than the Spanish gradiency ($M = -1.59$, $SD = 0.30$), $t(55) = 3.23$, $BF = 14.26$. Also, there was decisive evidence that Spanish consistency ($M = -17.90$, $SD = 3.51$) was higher than English consistency ($M = -21.35$, $SD = 6.09$), $t(55) = -4.58$, $BF = 737.64$. Although listeners were less consistent in English than in Spanish and had flatter VAS curves in general, we believe that our gradiency measure (which adjusts for consistency) is a valid index of gradiency. First, as shown in Figure *3.6*, one participant may exhibit lower gradiency (i.e., a steeper slope) and another higher gradiency (i.e., a shallower slope), even though both show relatively narrow response ranges and more variable ratings in the English VAS, compared to the Spanish one. Second, as illustrated in Figure *3.4*b, if participants had only one category in mind (treating all English sounds as Spanish /p/), their mean response to the 1 ms token would resemble the value in Figure *3.4*a (~70). This is not the case: the VAS rating is substantially lower (~30) than that, and all response curves in the analysis show increasing trends. Therefore, although their VAS functions are less sharp, participants nonetheless maintain mental representations of both English /b/ and /p/ categories, demonstrating that the VAS slope is a valid index of gradiency.

Participants demonstrated a substantial range of English proficiency. The average score on the picture-naming task was 77% (*SD* = 16%), with scores from 34% to 100%. We examined whether English proficiency modulated the overall response range. Specifically, we tested whether higher proficiency was associated with sharper category representations, reflected by a wider response range (i.e., a pattern more similar to that observed in L1). The VAS data for the two proficiency groups (based on a median split) are presented in Figure *3.7*, which shows minimal differences between high- and low-proficiency participants (range: high proficiency group, *M* = 56.64, *SD* = 15.25; low proficiency group, *M* = 53.34, *SD* = 14.25). Consistent with this visual impression, the correlation between English proficiency and VAS response range was negligible, $\rho = -.04$, *BF* = 0.31. To further evaluate these effects, we conducted independent-samples t-tests comparing response ranges between the two proficiency groups. The group comparison yielded moderate evidence for the null, $t(53.90) = 0.47$, *BF* = 0.30.

Figure 3.7
Average English VAS Data for (a) Low English Proficiency Group and (b) High English Proficiency Group



Note. Each light-colored curve represents the average of a participant's response at each VOT value. The solid dark line is the fitting curve based on the rotated logistic function, while the dashed dark line connects the grand average of all participants' responses. The two proficiency groups were defined using a median split.

In summary, proficiency is not associated with the robustness of category representations in L2 English. Taken together with the cross-linguistic comparison presented above, these results indicate that across the range of proficiency, English /b/ and /p/ remain less robustly represented than their L1 counterparts.

3.3.2 Preliminary Analyses: VWP

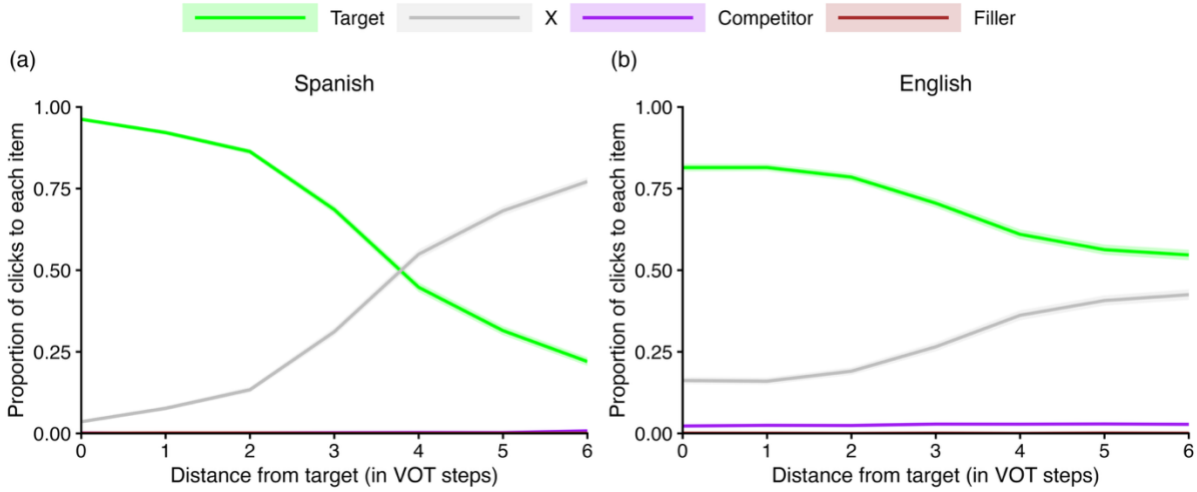
Then, we conducted group-level analyses of accuracy, RTs, and fixations in the lexical garden-path tasks for both Spanish and English, following the approach of previous studies (Kapnoula et al., 2021; McMurray et al., 2009). This step served to validate the VWP experiments.

The VOT step of the initial phoneme in each auditory stimulus was recoded to represent its acoustic distance from the target (henceforth, *tDist*), ranging from 0 to 6. To illustrate, consider a stimulus at VOT Step 1 (*VOT* = −35 ms in Spanish and 0 ms in English). In this case, tDist was coded as 0 for the extreme /b/ targets (e.g., *balanza* and *beachball*) and 6 for the extreme /p/ targets (e.g., *balacio* and *beachpit*). Likewise, for a stimulus at VOT Step 7 (VOT = +15 ms in Spanish and +48 ms in English), tDist was coded as 0 for the extreme /p/ targets (e.g., *palacio* and *peachpit*) and 6 for the extreme /b/ targets (e.g., *palanza* and *peachball*). This recoding allowed us to examine the effect of acoustic distance from the target across both voiced and voiceless stimuli.

First, we focus on the clicking response patterns. Figure *3.8* shows the clicking response patterns on the VWP tasks. For the Spanish VWP, in the case of completely unambiguous target stimuli (i.e., tDist = 0), average accuracy (i.e., proportion of trials in which participants clicked on the target picture) was 96.3% (*SD* = 7.1%). Participants clicked on the competitor, filler item, and X in 0.1% (*SD* = 0.4%), 0.1% (*SD* = 0.3%), and 3.5% (*SD* = 7.1%) of the trials, respectively. Response times averaged 1,452 ms (*SD* = 138 ms). For the

English VWP, mean accuracy for completely unambiguous target stimuli was high but noticeably lower (M = 81.4%, SD = 11.4%). Participants clicked on the competitor, filler item, and X in 2.3% (SD = 3.3%), 0.1% (SD = 0.3%), and 16.2% (SD = 12.0%) of the trials, respectively. Participants exhibited slightly slower RTs in English (M = 1,645 ms, SD = 195 ms). The shallow slope of the mouse-click responses across tDist in L1 Spanish (Figure *3.8*a) indicates that listeners generally exhibit gradient processing at the lexical level, which corroborates previous findings in L1 English (e.g., Andruski et al., 1994; McMurray et al., 2002; Utman et al., 2000).

Figure 3.8
Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in (a) Spanish VWP and (b) English VWP



Note. Shaded ribbons indicate standard errors of the mean (SEMs).

Importantly, Figure *3.8* provides further evidence of differences between L1 and L2 processing. Specifically, the effect of tDist differed between L1 and L2. Overall, as tDist increased, participants were less likely to click on the target, but this pattern was more pronounced for L1 Spanish than for L2 English. For Spanish, when the VOT was highly misleading (tDist = 6), participants selected the target in 21.6% of trials and the X in 77.6% of trials. This pattern closely resembled that observed when L1 English speakers performed the English VWP task (see Figure 5 in Kapnoula et al., 2021). In contrast, for highly

misleading items in L2 English, participants selected the target in 54.7% of trials, which was much higher than in L1, $t(62) = 8.98$, $BF = \infty$, providing decisive evidence for this difference. Similarly, they selected the X in only 42.5% of trials, which was much lower than in L1, $t(62) = -9.06$, $BF = \infty$, again providing decisive evidence. Hence, participants' clicking responses in L2 English were more influenced by disambiguating lexical information.

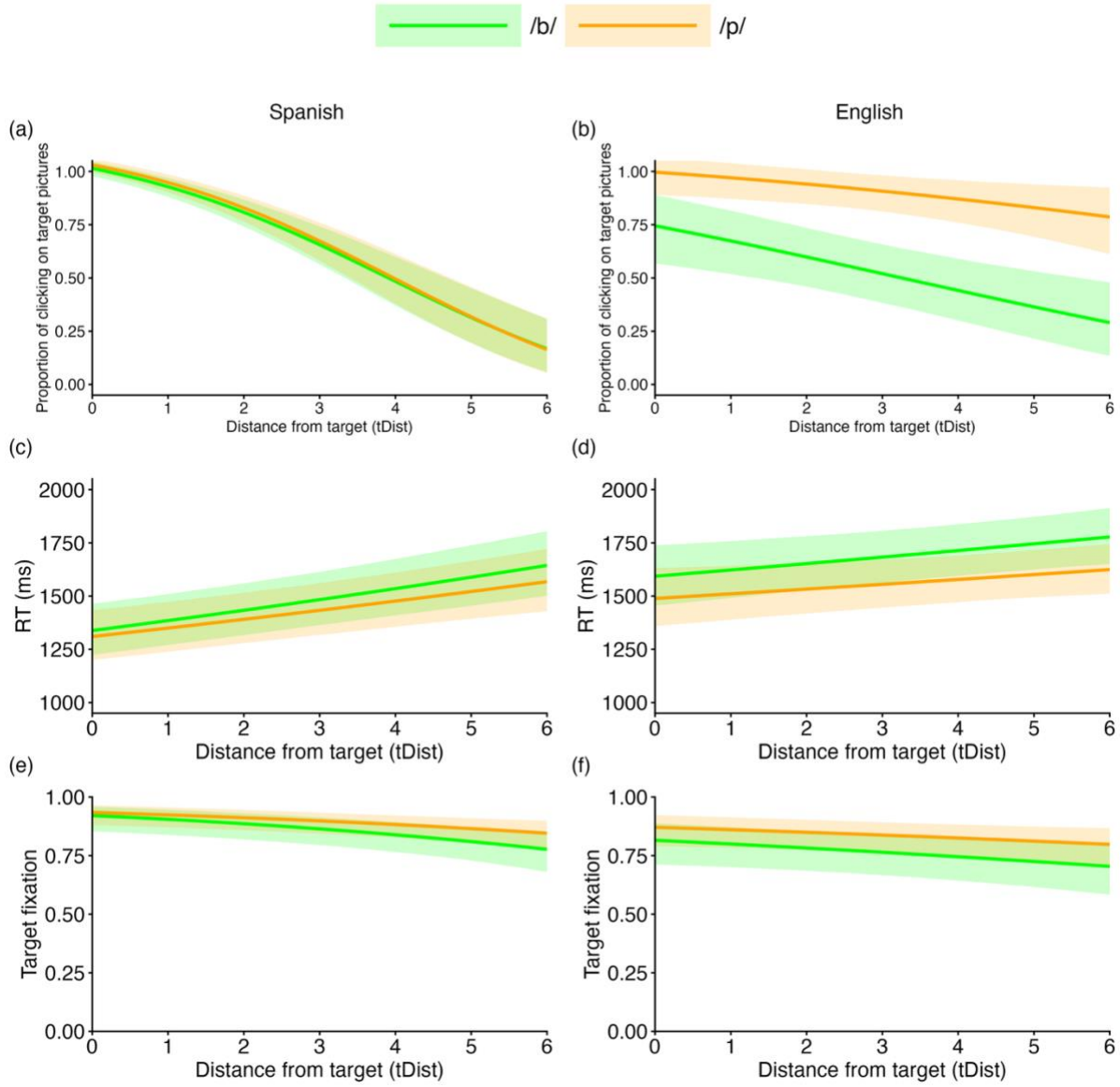
We statistically evaluated predictors of accuracy using three fixed effects. The first was *tDist*. The second was *splice*, indicating whether the onset and offset of a stimulus came from the same or a different item. The final predictor was *target voicing*, indicating whether the target began with /b/ (e.g., *balanza* in Spanish or *beachball* in English) or /p/ (e.g., *palacio* in Spanish or *peachpit* in English). Splice and target voicing were effect-coded (matched splice = 1; mismatched splice = −1; /b/ target = 1; /p/ target = −1), whereas tDist was centered. The dependent variable was accuracy (empirical-logit transformed), calculated as the proportion of trials in which participants clicked on the target picture.

We found decisive evidence for a negative effect of tDist in Spanish, $\beta = -0.54$, 95% *CrI* [−0.61, −0.46], $BF = \infty$, and in English, $\beta = -0.20$, 95% *CrI* [−0.30, −0.11], $BF =$ 1,999.00, indicating that listeners were more likely to select the target at lower tDist values in both languages (Figure *3.9*a–b). Decisive evidence was also observed for a positive effect of splice in Spanish, $\beta = 0.09$, 95% *CrI* [0.07, 0.11], $BF = \infty$, with more target responses when coarticulation matched the target. For English, we found decisive evidence for a negative effect of target voicing, $\beta = -0.68$, 95% *CrI* [−0.95, −0.41], $BF =$ 2,665.67, indicating more target responses for /p/-initial words across tDist values. This may reflect the similarity in VOT between English /b/ and Spanish /p/, leading to confusion for English /b/-initial targets; consequently, English /b/-onset words were harder than /p/-onset words. Importantly, in the same English VWP task with English monolinguals (Kapnoula et al., 2021), this target-voicing effect was absent, suggesting it is specific to L2 processing in Spanish speakers. For

Spanish, we found very strong evidence for an interaction between tDist and splice, $\beta = 0.01$, 95% *CrI* [0.00, 0.02], *BF* = 56.97, such that the splice effect (i.e., more target responses when the splice was matched) was larger at higher tDist values. Complete results are reported in Tables A3.2 and A3.3.

Figure 3.9

Predicted Accuracy, RT, and Proportion of Fixations on the Target as a Function of tDist for Voiced and Unvoiced Targets



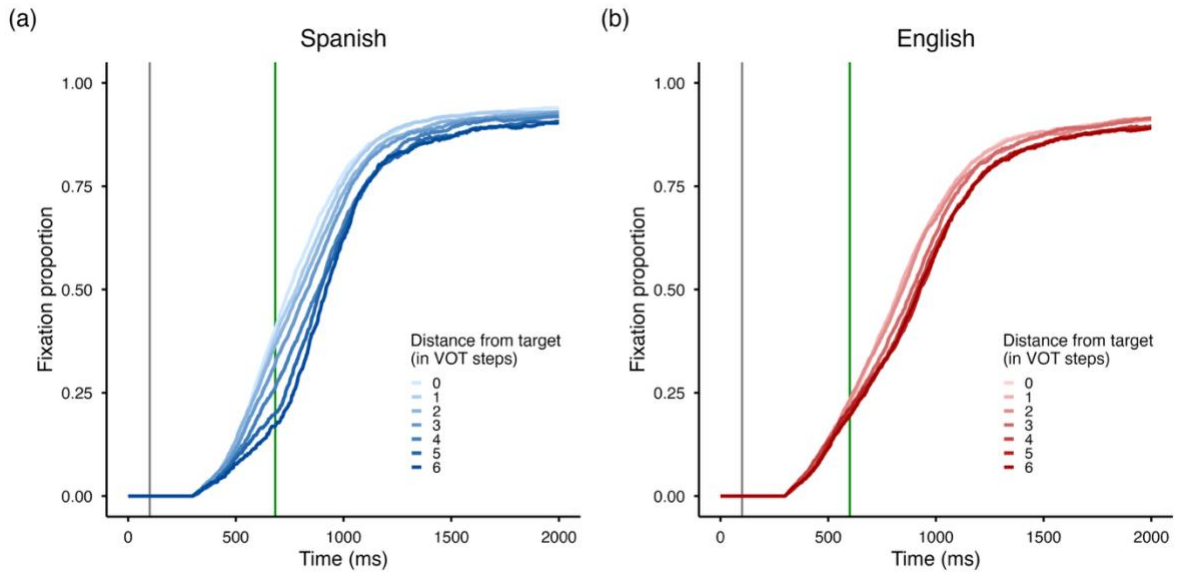
Note. (a, b) Predicted accuracy as a function of tDist for voiced and unvoiced targets; (c, d) predicted RT as a function of tDist for voiced and unvoiced targets; (e, f) predicted proportion of fixations on the target (from POD to trial end) as a function of tDist for voiced and unvoiced targets. Across panels, the left column displays the data for Spanish VWP, while the right column shows the data for English VWP. Values on both axes were back-transformed from the model's transformed scale to their original scale for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

The second analysis focused on RT. The same random and fixed effects were applied. RTs exceeding ±3 SDs from the mean were treated as outliers, and the corresponding trials were removed from each VWP dataset. In addition, only trials in which participants ultimately clicked on the target picture were included. RTs were log-transformed, and tDist was centered.

We found decisive evidence supporting a positive effect of tDist in Spanish, $\beta = 0.03$, 95% *CrI* [0.02, 0.04], $BF = \infty$, and in English, $\beta = 0.02$, 95% *CrI* [0.01, 0.03], $BF = 332.33$, indicating longer RTs with higher tDist for both languages (Figure *3.9*c–d). Decisive evidence for a negative effect of splice was observed only for Spanish, $\beta = -0.02$, 95% *CrI* [−0.03, −0.02], $BF = \infty$, suggesting that participants responded faster when the coarticulation matched the target. For English, we found decisive evidence for an interaction between splice and target voicing, $\beta = 0.01$, 95% *CrI* [0.00, 0.01], $BF = 55.33$, indicating that the target-voicing effect (i.e., higher RTs when the word onset was /b/ compared with /p/) was larger when the splice was matched. Complete results are reported in Tables A3.4 and A3.5.

Next, we examined the proportion of time that participants fixated on the target. Similar to the RT analysis, we included only trials in which participants ultimately clicked on the target. As shown in Figure *3.9*e–f and Figure *3.10*, participants tended to look more at the target for smaller tDist values, with delayed and reduced fixations as tDist increased in both the English and Spanish VWPs. To statistically test these observations, we employed a Bayesian linear mixed-effects model with the same fixed and random effects as before for each language. The dependent variable was target fixations, defined as the average proportion of fixations on the target within the time window starting from the item POD, corrected for a 200-ms oculomotor delay, and extending until 2,000 ms. Target-fixation values were empirical-logit transformed, and tDist was centered.

Figure 3.10
Target Fixations as a Function of Time and tDist in (a) Spanish and (b) English VWPs



Note. The gray vertical line marks the 100 ms pre-stimulus silence, and the green line marks the point of disambiguation (POD), corrected for a 200 ms oculomotor delay.

We found decisive evidence for a negative effect of tDist in both Spanish, $\beta = -0.18$, 95% *CrI* [−0.25, −0.10], *BF* = 3,999.00, and English, $\beta = -0.10$, 95% *CrI* [−0.16, −0.03], *BF* = 130.15), corroborating that there were fewer fixations on the target as tDist increased (Figure *3.9*e–f). Decisive evidence for a positive effect of splice was observed for Spanish, $\beta = 0.16$, 95% *CrI* [0.12, 0.21], *BF* = ∞, suggesting that participants fixated on the target more when the coarticulation matched the target. In addition, we found decisive evidence for an interaction between splice and target voicing in Spanish, $\beta = -0.09$, 95% *CrI* [−0.13, −0.04], *BF* = 3,999.00, indicating that participants were more likely to fixate on the target picture when the word began with /p/ in the splice-matching condition only. Complete results are reported in Tables A3.6 and A3.7.

Figure *3.10* also provides a third piece of evidence for differences in processing between L1 and L2. Specifically, participants exhibited a less gradient target-fixation pattern in L2 English (Figure *3.10*b) than in L1 Spanish (Figure *3.10*a). To statistically test whether target-fixation patterns differed across tDist between L1 Spanish and L2 English, we

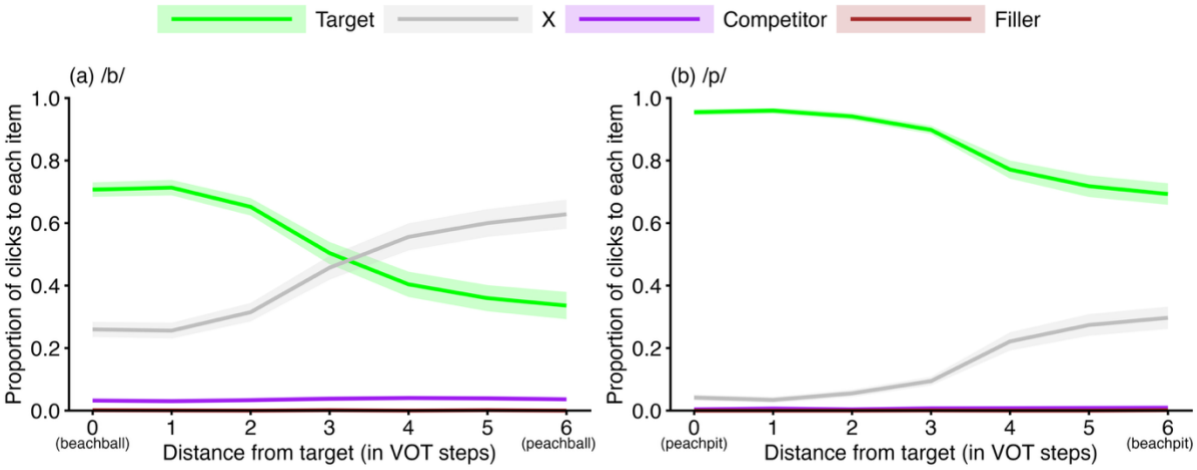
combined both datasets and reran the same analyses, adding language (1 = English; −1 = Spanish) as an additional predictor. The complete results are reported in Table A3.8. Notably, a tDist × language interaction on fixation showed very strong Bayesian evidence, $\beta = 0.05$, 95% *CrI* [0.01, 0.09], $BF = 38.80$, confirming that the observed difference between L1 and L2 fixation patterns was statistically reliable.

Thus, participants showed differences between L1 and L2 processing across multiple measures: VAS responses, VWP clicking responses, and VWP fixation patterns. The latter two, in particular, reflected a general tendency for listeners to rely more on lexical and less on acoustic-phonetic information in L2, possibly due to less robust category representations.

Next, we explore how proficiency modulates click responses to the target picture in L2 English. Figure *3.11* shows the clicking-response patterns in the English VWP for /b/-onset and /p/-onset words. Figures 3.12–3.14 illustrate the patterns for listeners with higher and lower English proficiency (based on a median split): Figure *3.12* presents the overall pattern, Figure *3.13* the /b/-onset words, and Figure *3.14* the /p/-onset words. Visually, listeners with higher proficiency appeared to show a steeper decrease in target-picture clicks across tDist compared to those with lower proficiency. To test this statistically, we repeated the analyses described above, replacing splice with English proficiency. The model therefore included three predictors: tDist (centered), English proficiency (centered), and target voicing (effect-coded). The dependent variable was accuracy (empirical-logit transformed). If listeners with higher English proficiency relied less on top-down information, we would expect a two-way interaction between tDist and English proficiency, such that higher-proficiency listeners would click on the target less when tDist was high.

Figure 3.11

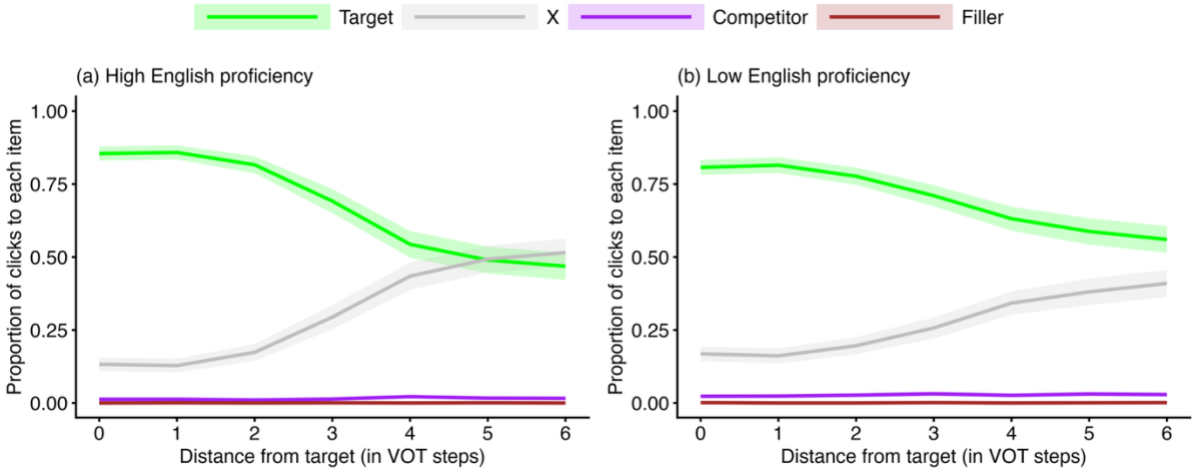
Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) /b/-Onset Words and (b) /p/-Onset Words



Note. Shaded ribbons indicate standard errors of the mean (SEMs).

Figure 3.12

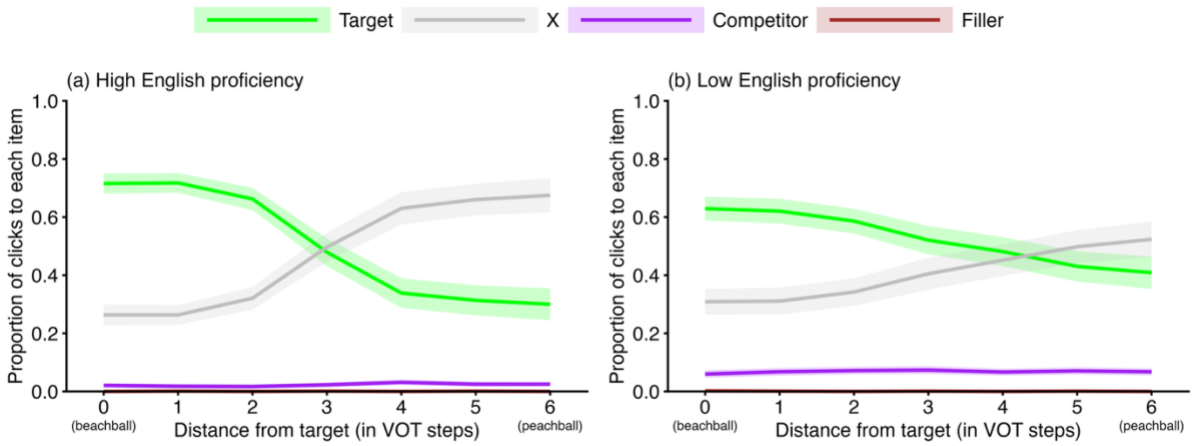
Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency



Note. Shaded ribbons indicate standard errors of the mean (SEMs). The high and low proficiency is based on a median split.

Figure 3.13

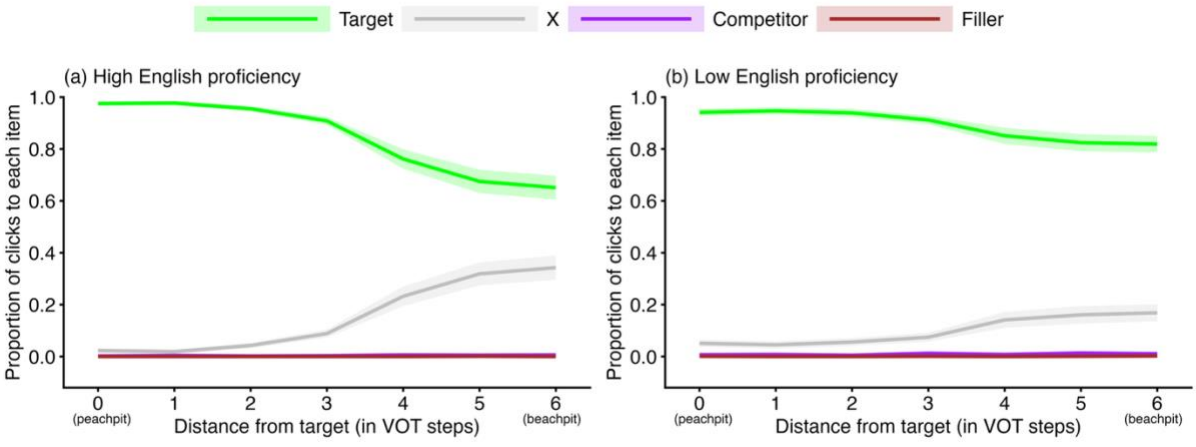
Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency for /b/-Onset Words



Note. Shaded ribbons indicate standard errors of the mean (SEMs). The high and low proficiency is based on a median split.

Figure 3.14

Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency for /p/-Onset Words



Note. Shaded ribbons indicate standard errors of the mean (SEMs). The high and low proficiency is based on a median split.

Indeed, we found decisive evidence for an interaction between tDist and English proficiency on accuracy, $\beta = -0.01$, 95% *CrI* [−0.01, −0.01], $BF = \infty$, indicating that listeners with higher proficiency clicked on the target less when tDist was high (Figure A3.1). Although the three-way interaction between tDist, English proficiency, and target voicing showed strong evidence, $\beta = 0.00$, 95% *CrI* [−0.00, 0.00], $BF = 15.91$, the 95% credible interval overlapped with zero, suggesting that the effect may not be robust. The direction of the effect indicated that listeners with higher proficiency showed a steeper decrease across tDist more clearly for /b/-onset words than for /p/-onset words (Figure A3.2). Recall that a steeper decrease in target clicking across tDist is considered the optimal pattern, as seen in the L1 Spanish data (Figure *3.9*a). Taken together, these results suggest that the relationship between English proficiency and tDist on clicking behavior is not clearly modulated by target voicing. Higher-proficiency listeners may generally rely less on top-down information compared to lower-proficiency listeners. Complete results are reported in Table A3.9.

In summary, the preliminary analyses revealed a strong effect of tDist across all measures in both the Spanish and English VWPs. Participants were more likely to select the target, responded more quickly, and fixated on the target more frequently when the acoustic distance from the target was small. These findings align with those of Kapnoula et al. (2021), confirming that participants performed the task appropriately and that the manipulation was effective. Notably, in the English VWP, we found compelling evidence that listeners provided more target responses for /p/-initial words across tDist, suggesting that English words beginning with /b/ were more challenging for L1 Spanish speakers than those beginning with /p/. Furthermore, participants exhibited a less gradient target-fixation pattern in English than in Spanish, reinforcing the VAS and click-response findings that L2 listeners tend to rely more on lexical information and less on acoustic-phonetic cues. Finally, within the L2, reliance on top-down information appeared to decrease with proficiency.

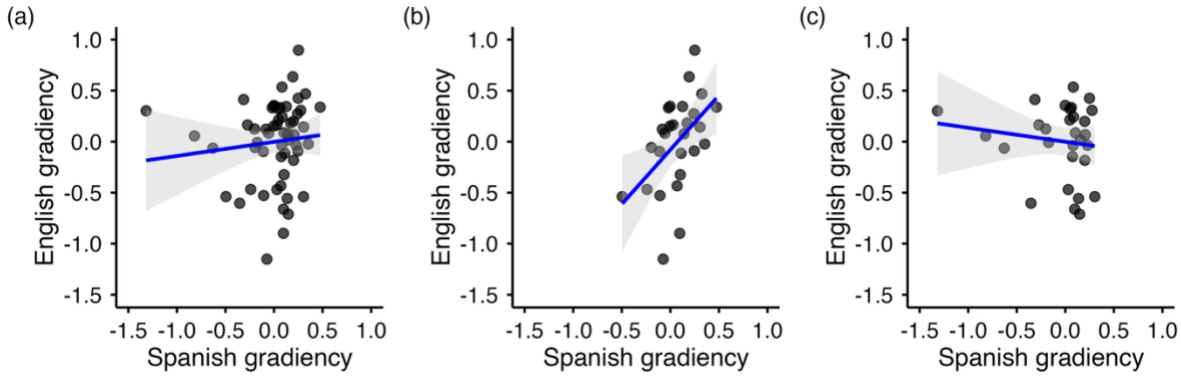
3.3.3 Is L2 Gradiency/Consistency Related to L1 Gradiency/Consistency?

We first report tests examining whether L2 gradiency/consistency is related to L1 gradiency/consistency. Spearman's correlations were conducted among L1 gradiency, L2 gradiency, L2 proficiency (English picture naming), and the covariates (i.e., working memory, inhibitory control, English listening exposure, and musical training duration), as well as demographic variables. All data from the spatial Stroop and Corsi tasks were retained, except for two spatial Stroop trials with RTs exceeding 2,000 ms, which were excluded. Table A3.10 presents the correlation results for all variables. Here, we focus only on our main research interests, namely, whether L1 gradiency is related to L2 gradiency and whether L1 consistency is related to L2 consistency. Other statistically meaningful findings will also be reported.

The correlation between Spanish and English gradiency was weak, $\rho = .18$, $BF = 0.39$, indicating weak evidence in favor of the null hypothesis, that is, no meaningful relationship between L1 and L2 gradiency (Figure *3.15*a). Recall that the preliminary results above suggested that listeners with low English proficiency relied less on bottom-up acoustic-phonetic information, pointing to weaker encoding and/or noisier processing of acoustic cues. However, listeners with different L2 proficiency levels did not differ in the robustness of their category representations of English bilabial stops. Here, we test whether listeners with varying proficiency show L2 gradiency that is more closely aligned with their L1 gradiency. To address this issue, we divided participants into high- and low-proficiency groups (using a median split) and reanalyzed the data. For the high proficiency group, the correlation between L1 and L2 gradiency was positive, $\rho = .53$, $BF = 6.79$, with moderate evidence (Figure *3.15*b). In contrast, the low proficiency group showed a weak correlation, $\rho = -.12$, $BF = 0.50$, offering weak evidence in favor of the null (Figure *3.15*c). These results suggest that L1 and L2 gradiency are positively linked, but only among listeners with higher L2 proficiency.

Taken together with the proficiency-related differences in mouse-click responses, these findings underscore the importance of considering L2 proficiency. The distinction between L1 and L2 processing is not absolute; with sufficient proficiency, L2 processing can increasingly resemble that of the L1. Moreover, because response range did not differ across proficiency groups, as mentioned above, the differing L1–L2 correlations are unlikely to reflect confounding factors and instead point to a genuine difference in fine-grained acoustic processing.

Figure 3.15
Correlation between L1 and L2 Gradiency in (a) the Full Dataset, (b) High Proficiency Listeners, and (c) Low Proficiency Listeners

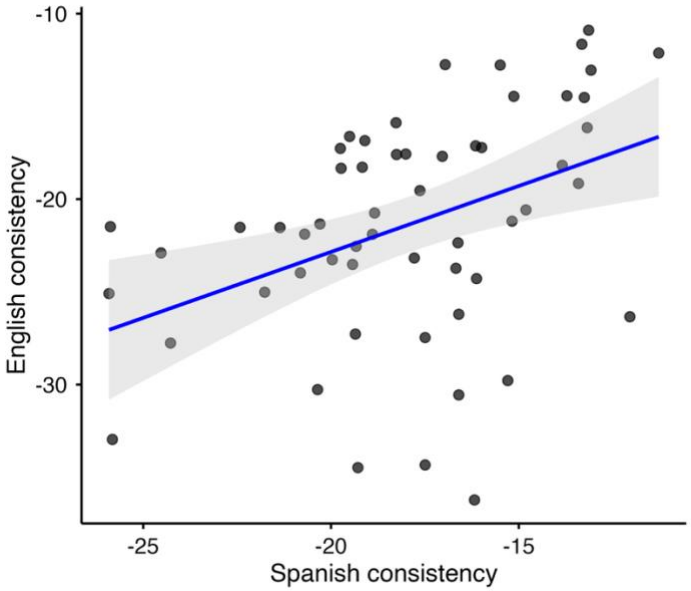


Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

Regarding the relationship between L1 and L2 consistency, we found very strong evidence for a positive correlation between Spanish consistency and English consistency, ρ = .44, BF = 27.18 (Figure *3.16*). This suggests that consistency may represent a more general trait than gradiency. The only other meaningful finding was a strong positive correlation between Spanish gradiency and English consistency, ρ = .40, BF = 12.86, indicating that listeners with greater gradiency in Spanish exhibited higher consistency in English (Figure A3.3). This pattern is consistent with the possibility that greater L1 gradiency supports the acquisition of fine-grained acoustic-phonetic cues in the L2, which may in turn be associated

with higher consistency. Alternatively, more stable phonetic categorization in the L2 could facilitate gradiency in the L1.

Figure 3.16
Correlation between L1 Spanish Consistency and L2 English Consistency



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

The relationship between L2 gradiency and L2 proficiency is weak, with moderate evidence for the null, $\rho = .02$, $BF = 0.31$, while the relationship between L2 consistency and L2 proficiency is inconclusive, $\rho = .21$, $BF = 1.14$. These findings provide no clear evidence for a relationship between L2 proficiency and L2 gradiency or consistency.

3.3.4 Gradiency Effects on Initial Lexical Activation and Speech Perception Flexibility in L1 and L2

Our first main research question asks whether gradiency predicts initial lexical activation and speech perception flexibility in L1 Spanish and L2 English. Initial lexical activation was indexed by participants' fixations on /b/- and /p/-onset pictures as a function of VOT prior to the POD. Accordingly, this measure captures early lexical activation based on phoneme onset, independent of word offset timing or participants' final responses.

We ran Bayesian linear mixed-effects analyses to examine the relationship between VAS gradiency (the inverse of slope) and initial lexical activation (i.e., proportion of fixations on /b/ and /p/-onset pictures). Each trial was coded as 1 if the participant fixated on the /b/ or /p/-onset picture between stimulus onset and the POD, both adjusted for a 200 ms oculomotor delay; otherwise, it was coded as 0. The binary values were then averaged for each participant and each unique combination of VOT step and gradiency (i.e., within each condition cell). The resulting proportions (i.e., the number of trials divided by the total of six trials per condition cell) were empirical-logit transformed and analyzed as a function of VOT step (centered) and gradiency (centered).

We fitted two models for each language: one with the proportion of fixations on /b/-onset pictures as the dependent variable and one with the proportion of fixations on /p/-onset pictures as the dependent variable. The models included by-subject and by-item random intercepts, as well as random slopes for VOT step. Because lower VOT values correspond to /b/ and higher VOT values correspond to /p/ in both Spanish and English, stronger initial lexical activation was reflected in greater proportions of fixations on /b/-onset pictures at lower VOT values and greater proportions of fixations on /p/-onset pictures at higher VOT values. Critically, this pattern was indexed by the interaction between VOT step and gradiency, rather than by a main effect of gradiency alone. This approach allows us to assess the effects of gradiency on lexical activation independently of potential differences in overall looking patterns or other biases, and separately for /b/- and /p/-onset items. Details of the statistical framework are provided in Section 2.2.6, and the specific model formulas and statistical results are reported in Appendix 4.

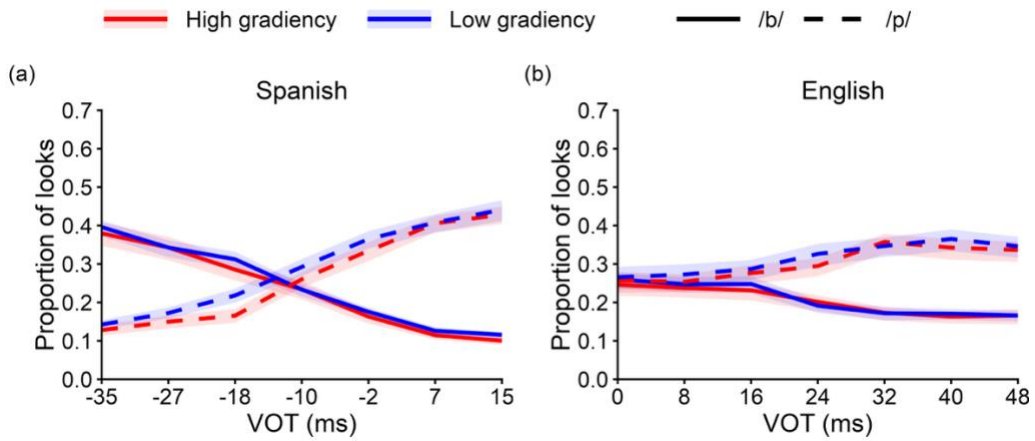
Figure *3.17* illustrates how the proportion of fixations on /b/-onset and /p/-onset pictures prior to the POD varied as a function of VOT and gradiency in Spanish and English, respectively. As expected, listeners showed fewer fixations on /b/-onset pictures and more

fixations on /p/-onset pictures as VOT increased. Gradiency seems not affect the VOT-compatible looking patterns. This observation was confirmed by the statistical analyses reported below.

Figure *3.18* shows the model-predicted proportion of fixations on /b/-onset or /p/-onset pictures as a function of VOT and gradiency for both Spanish and English. First, we report the analyses of the proportion of fixations on /b/-onset pictures in Spanish. As expected, there was decisive evidence for a main effect of VOT step, $\beta = -0.24$, 95% *CrI* [−0.32, −0.17], $BF = \infty$, reflecting fewer fixations on /b/-onset pictures as VOT increased overall. We also found weak evidence for the null effect of gradiency, $\beta = 0.10$, 95% *CrI* [−0.26, 0.47], $BF = 0.79$, indicating that there was no robust evidence that listeners with different levels of gradiency differed in their overall tendency to look at /b/-onset pictures. If more gradient listeners exhibit stronger initial lexical activation, we would expect an interaction between VOT step and gradiency, such that listeners with higher gradiency show more fixations on /b/-onset pictures at lower VOT values than listeners with lower gradiency. However, we did not find evidence for such an interaction, $\beta = -0.02$, 95% *CrI* [−0.10, 0.07], $BF = 0.39$ (Figure *3.18*a). This suggests that gradiency does not facilitate initial lexical activation. Complete results are presented in Table A3.11.

Figure 3.17

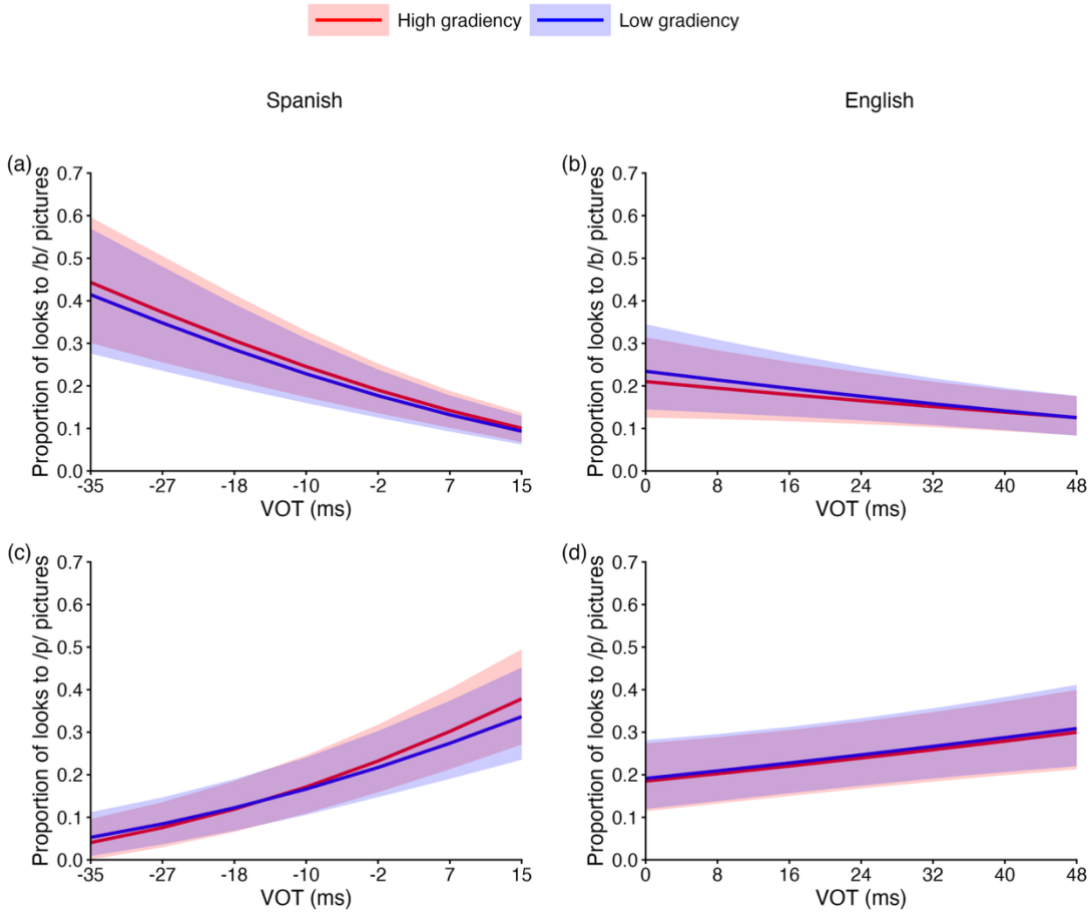
Proportion of Fixations on /b/-Onset Pictures and /p/-Onset Pictures as a Function of VOT and Gradiency in (a) Spanish and (b) English



Note. "Low gradiency" and "high gradiency" were defined by a median split. "/b/" and "/p/" refer to fixations on pictures whose target words began with /b/ and /p/ before the point of disambiguation (POD), respectively. Shaded ribbons indicate ±1 standard error of the mean (SE) around each curve, computed across participants at each VOT.

Figure 3.18

(a,b) Proportion of Fixations on /b/-Onset Pictures and (c,d) Proportion of Fixations on /p/-Onset Pictures as a Function of VOT and Gradiency for Spanish (Left Column) and English (Right Column)



Note. Lines represent model-predicted values. "Low gradiency" represents one standard deviation below the mean, and "high gradiency" represents one standard deviation above the mean. These splits are provided for illustrative purposes only, as gradiency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

Next, we report the analyses of the proportion of fixations on /b/-onset pictures in English. There was very strong evidence for a main effect of VOT step, $\beta = -0.08$, 95% *CrI* [−0.14, −0.02], $BF = 93.12$, reflecting fewer fixations on /b/-onset pictures as VOT increased overall. The effect of gradiency was negligible, with moderate evidence supporting the null, $\beta = -0.04$, 95% *CrI* [−0.36, 0.27], $BF = 0.27$. Evidence for the interaction between VOT step

and gradiency was inconclusive, $\beta = 0.02$, 95% *CrI* [−0.03, 0.08], $BF = 1.32$, indicating that the data did not provide clear support for either the presence or absence of an interaction (Figure *3.18*b). Complete results are presented in Table A3.12.

For the proportion of fixations on /p/-onset pictures in Spanish, there was decisive evidence for a main effect of VOT step, $\beta = 0.26$, 95% *CrI* [0.19, 0.34], $BF = \infty$, reflecting that listeners looked more to /p/-onset pictures as VOT increased overall. The effect of gradiency was inconclusive, $\beta = 0.11$, 95% *CrI* [−0.24, 0.46], $BF = 0.86$. There was moderate evidence for an interaction between VOT step and gradiency, $\beta = 0.07$, 95% *CrI* [−0.02, 0.16], $BF = 6.95$; however, because the 95% CrI overlapped with 0, the effect was not considered robust (Figure *3.18*c; see Table A3.13 for complete results).

Turning to /p/-onset pictures in English, as expected, there was again decisive evidence for a main effect of VOT step, $\beta = 0.08$, 95% *CrI* [0.03, 0.13], $BF = 144.45$, reflecting an overall increase in fixations on /p/-onset pictures as VOT increased. There was moderate evidence supporting the null effect of gradiency, $\beta = -0.04$, 95% *CrI* [−0.48, 0.39], $BF = 0.18$. There was also very strong evidence for the null interaction between VOT step and gradiency, $\beta = 0.00$, 95% *CrI* [−0.06, 0.05], $BF = 0.02$ (Figure *3.18*d; see Table A3.14 for complete results).

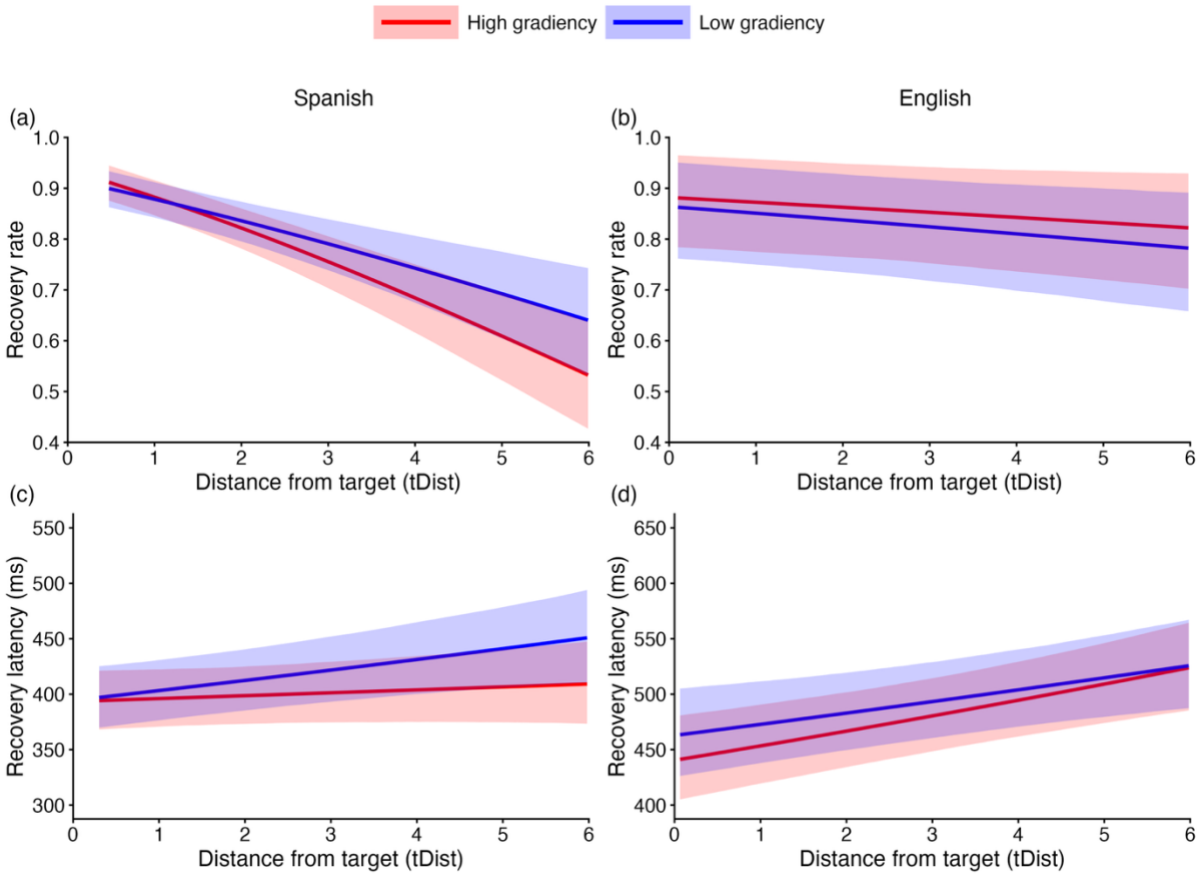
Next, we examined whether gradiency predicts speech perception flexibility. Speech perception flexibility was assessed using two measures: (1) recovery rate, defined as the proportion of garden-path trials in which participants ultimately fixated on the target after the point of disambiguation (POD), and (2) recovery latency, defined as the time taken to shift gaze to the target following the POD. Analyses were limited to garden-path trials, that is, trials in which participants initially fixated on a competitor (e.g., *peachpit* when hearing *peachball*) prior to the POD, because this fixation is a prerequisite for recovery. Higher recovery rates and faster recovery latencies indicate greater speech perception flexibility. In

the present analyses, we focus on acoustic distance from the target (tDist), gradiency, and their interaction.

Recovered trials were defined as cases in which participants initially fixated on the competitor picture before the POD and subsequently directed their gaze to the target after the POD. This analysis included all trials where participants clicked on any picture (including X) at the end. Each garden-pathed trial was coded as 1 if the participant fixated on the target at any point after the POD of that trial's stimulus (adjusted for the oculomotor delay); otherwise, it was coded as 0. These binary values were then averaged for each participant and each unique combination of tDist and gradiency (i.e., each condition cell). The resulting proportions (i.e., the number of recovered trials divided by the number of garden-pathed trials per cell) were empirical-logit transformed and served as the dependent variable in the analyses.

Figure *3.19* shows the model-predicted recovery rate and recovery latency as a function of tDist and gradiency for Spanish and English. There was decisive evidence for a negative relationship between tDist and recovery rate, $\beta = -0.18$, 95% *CrI* [−0.22, −0.13], *BF* = ∞. As expected, greater tDist was associated with a lower recovery rate. The effect of gradiency showed moderate evidence but overlapped with zero and was therefore not robust, $\beta = -0.24$, 95% *CrI* [−0.58, 0.10], *BF* = 5.33. If more gradient listeners exhibited greater speech perception flexibility, we would expect an interaction between tDist and gradiency, such that listeners with higher gradiency would show a higher recovery rate when tDist is high (Kapnoula et al., 2021). Although the Bayes factor (*BF* = 13.95) provided strong evidence for this interaction (but in the opposite direction), the 95% credible interval included zero, $\beta = -0.11$, 95% *CrI* [−0.23, 0.01], indicating uncertainty regarding the direction and strength of the effect (Figure *3.19*a). Full model results are reported in Table A3.15.

Figure 3.19
(a,b) Recovery Rate and (c,d) Recovery Latency as a Function of tDist and Gradiency for Spanish (Left Column) and English (Right Column)



Note. Lines represent model-predicted values. "Low gradiency" represents one standard deviation below the mean, and "high gradiency" represents one standard deviation above the mean. These splits are provided for illustrative purposes only, as gradiency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

For L2 English, there was very strong evidence for a negative effect of tDist on recovery rate, $\beta = -0.04$, 95% *CrI* [−0.07, −0.01], $BF = 35.70$, indicating that listeners exhibited lower recovery rates as tDist increased (Figure *3.19*b). As in L1 Spanish, gradiency did not predict recovery rate in L2 English. All effects involving gradiency were negligible (*BF*s = 0.53–1.91). Full results are reported in Table A3.16.

Next, we examined the relationship between gradiency and recovery latency (i.e., the time it took participants to recover from garden paths). Recovery latency was calculated as

the log-transformed time from the POD to the first fixation on the target. Only recovered trials were included in this analysis. Bayesian mixed-effects models were employed, incorporating the same fixed effects as in the recovery rate analyses.

For L1 Spanish, we found very strong evidence for a main effect of tDist, $\beta = 0.01$, 95% *CrI* [0.00, 0.03], $BF = 31.00$, indicating that listeners recovered more slowly as tDist increased. In addition, listeners with higher gradiency recovered faster overall, with very strong evidence for a main effect of gradiency, $\beta = -0.09$, 95% *CrI* [−0.17, −0.01], $BF = 38.22$. If more gradient listeners exhibited higher speech perception flexibility, we would expect an interaction between tDist and gradiency on recovery latency, indicating that listeners with higher gradiency recover faster when tDist is high. Indeed, we found strong evidence for this interaction, $\beta = -0.03$, 95% *CrI* [−0.05, −0.00], $BF = 20.16$, suggesting that listeners with higher gradiency recovered more quickly when tDist was high (Figure *3.19*c). This finding aligns with previous evidence showing that listeners with higher gradiency exhibit greater speech perception flexibility in L1 English (Kapnoula et al., 2021). Complete results are reported in Table A3.17.

In contrast to the interaction between tDist and gradiency observed in L1 Spanish, all effects involving gradiency were inconclusive for L2 English (*BF*s = 1.56–1.75). The only notable effect was a main effect of tDist, indicating slower recovery at greater tDist, with decisive evidence, $\beta = 0.03$, 95% *CrI* [0.01, 0.04], $BF = 887.89$ (Figure *3.19*d). Full results are reported in Table A3.18.

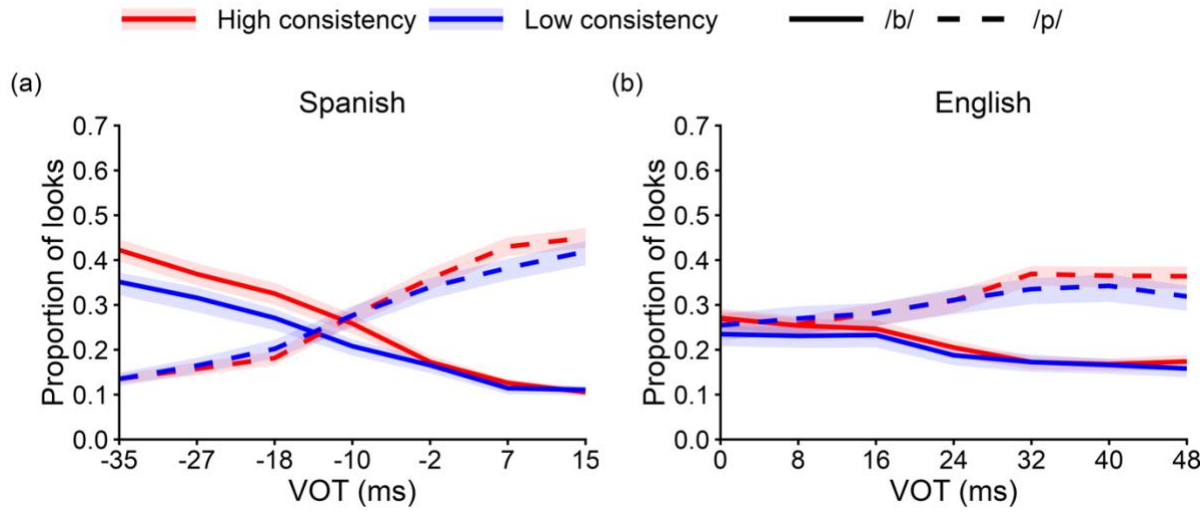
Based on the findings reported above, gradiency does not predict initial lexical activation across phoneme categories and languages. However, gradiency appears to play an important role in speech perception flexibility in L1 Spanish, but not in L2 English. This pattern is consistent with our previous suggestion that L2 listeners tend to rely less on bottom-up cues.

3.3.5 Consistency Effects on Initial Lexical Activation and Speech Perception Flexibility in L1 and L2

We examined whether consistency affects initial lexical activation and speech perception flexibility. All predictors were identical to those in the gradiency analyses above, except that the consistency measure (centered) was used in place of gradiency (or proficiency). We focus on the main effect and interactions involving consistency, as the effects of VOT step and tDist remained unchanged.

Figure *3.20*a and 3.20b illustrate how the proportion of fixations on /b/-onset and /p/-onset pictures prior to the POD varied as a function of VOT and consistency in Spanish and English, respectively. Listeners with higher consistency exhibited stronger VOT-compatible looking patterns, showing more fixations on /b/-onset pictures at lower VOT values and more fixations on /p/-onset pictures at higher VOT values. This pattern was confirmed by the statistical analyses reported below.

Figure 3.20
Proportion of Fixations on /b/-Onset Pictures and /p/-Onset Pictures as a Function of VOT and Consistency in (a) Spanish and (b) English

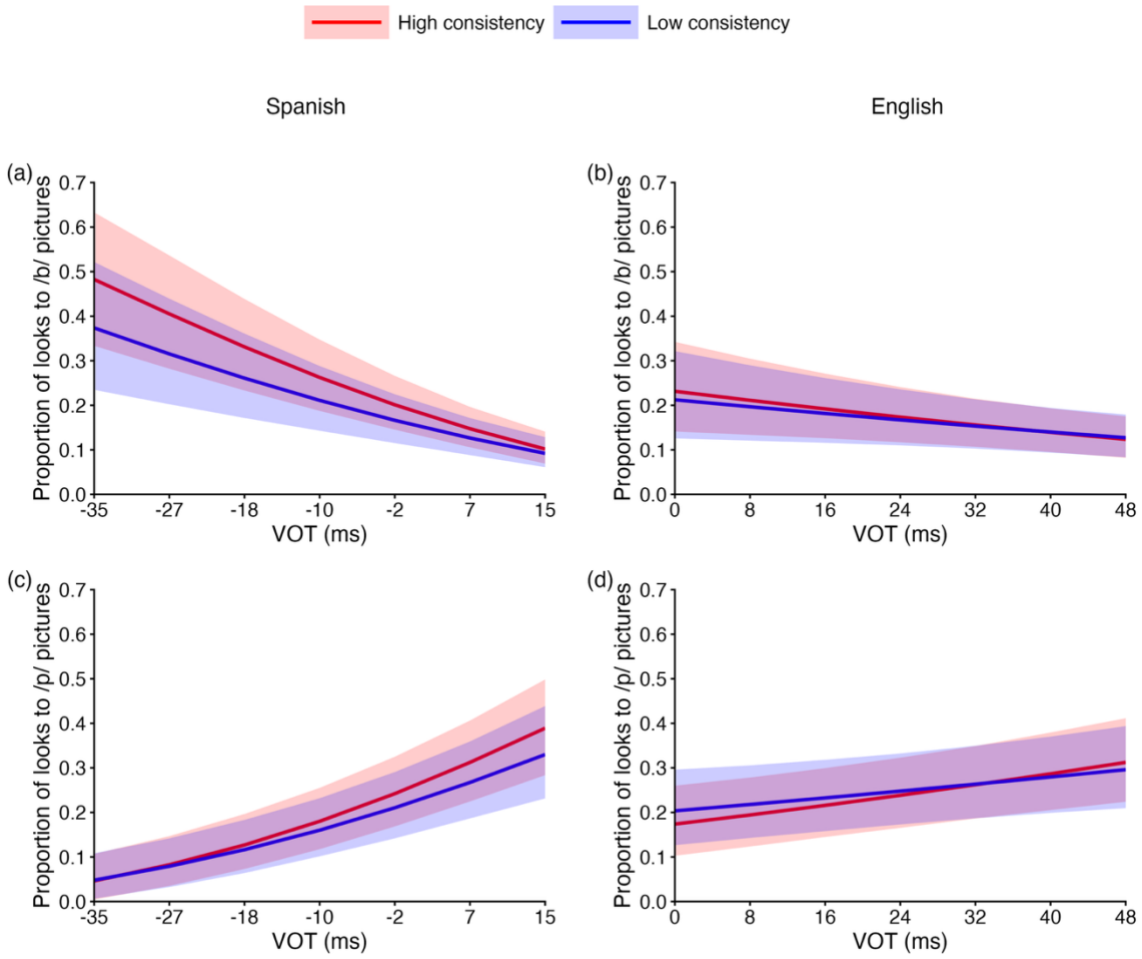


Note. "Low consistency" and "high consistency" were defined by a median split. "/b/" and "/p/" refer to fixations on pictures whose target words began with /b/ and /p/ before the point of disambiguation (POD), respectively. Shaded ribbons indicate ±1 standard error of the mean (SE) around each curve, computed across participants at each VOT.

Figure *3.21* shows the model-predicted proportion of fixations on /b/-onset or /p/-onset pictures as a function of VOT and consistency for both Spanish and English. First, we report the analyses of the proportion of fixations on /b/-onset pictures in Spanish. We found moderate evidence for a main effect of consistency, $\beta = 0.02$, 95% *CrI* [−0.01, 0.06], $BF = 6.14$, indicating that listeners with higher consistency showed a greater overall tendency to look at /b/-onset pictures. If more consistent listeners exhibit stronger initial lexical activation, we would expect an interaction between VOT step and consistency, such that listeners with higher consistency show more fixations on /b/-onset pictures at lower VOT values than listeners with lower consistency. Indeed, we found strong evidence for this interaction, $\beta = -0.01$, 95% *CrI* [−0.02, −0.00], $BF = 15.63$, indicating that listeners with higher consistency looked more to /b/-onset pictures when VOT was low (Figure *3.21*a). Complete results are presented in Table A3.19.

Figure 3.21

(a,b) Proportion of Fixations on /b/-Onset Pictures and (c,d) Proportion of Fixations on /p/-Onset Pictures as a Function of VOT and Consistency for Spanish (Left Column) and English (Right Column)



Note. Lines represent model-predicted values. "Low consistency" represents one standard deviation below the mean, and "high consistency" represents one standard deviation above the mean. These values are shown for illustrative purposes only, as consistency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale for ease of interpretation. Shaded ribbons indicate 95% credible intervals around the model-predicted values.

In contrast to the corresponding Spanish results, evidence for the interaction between VOT step and consistency on the proportion of fixations on /b/-onset pictures in L2 English was inconclusive, $\beta = 0.00$, 95% *CrI* [−0.01, 0.00], $BF = 1.32$, indicating that the data did not provide clear support for either the presence or absence of an interaction (Figure *3.21*b).

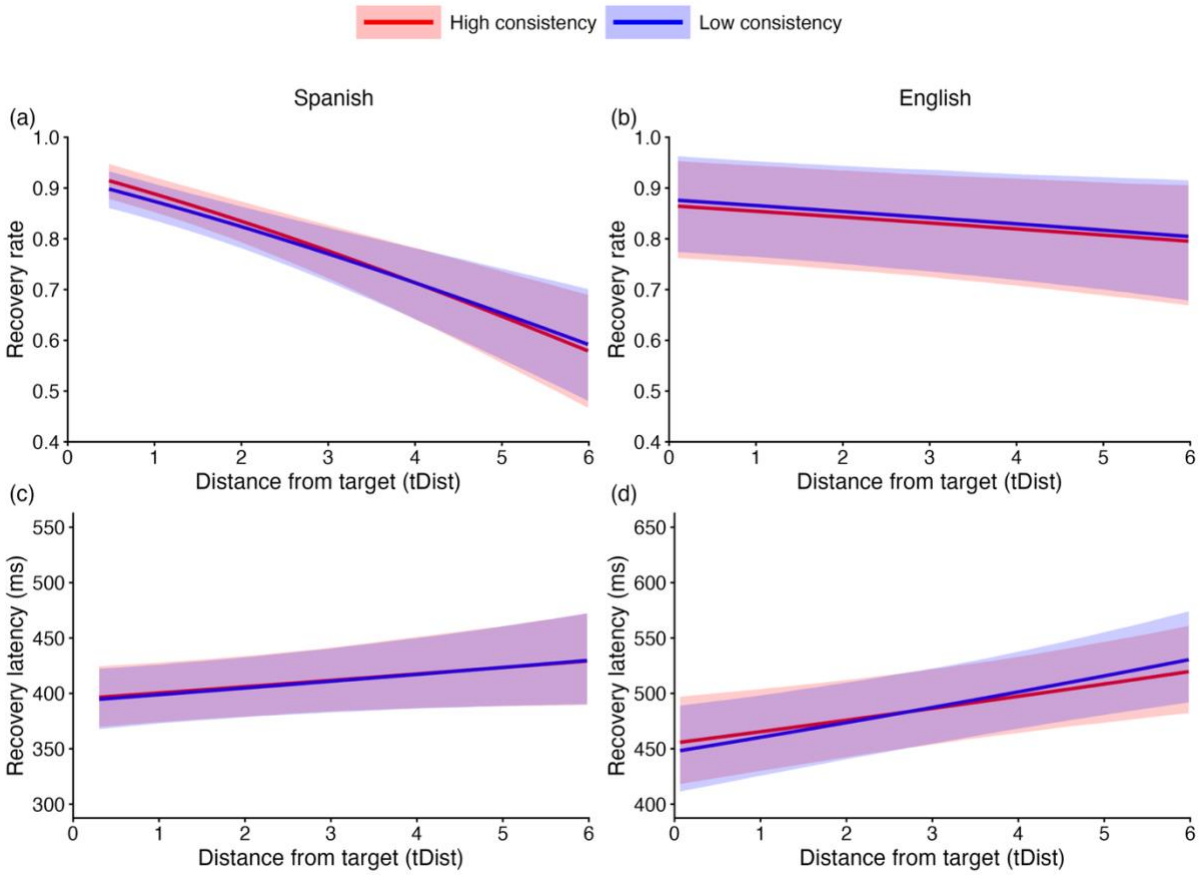
Evidence for a main effect of consistency was negligible (*BF* = 0.10). Complete results are presented in Table A3.20.

For the proportion of fixations on /p/-onset pictures in Spanish, there was weak evidence for a main effect of consistency, β = 0.02, 95% *CrI* [−0.01, 0.05], *BF* = 3.72. This effect was qualified by an interaction between VOT step and consistency, β = 0.01, 95% *CrI* [0.00, 0.01], *BF* = 4.42, with moderate evidence. This interaction indicates that listeners with higher consistency showed relatively more fixations on /p/-onset pictures at higher VOT values, compared to listeners with lower consistency (Figure *3.21*c), mirroring the pattern observed for fixations on /b/-onset pictures in Spanish (see Table A3.21 for complete results).

Turning to /p/-onset pictures in English, the main effect of consistency was negligible (*BF* = 0.03). Notably, there was moderate evidence for an interaction between VOT step and consistency, β = 0.00, 95% *CrI* [0.00, 0.01], *BF* = 8.21. This pattern indicates that listeners with higher consistency showed relatively more fixations on /p/-onset pictures at higher VOT values than listeners with lower consistency (Figure *3.21*d; see Table A3.22 for complete results).

Figure *3.22* shows the model-predicted recovery rate and recovery latency as a function of tDist and consistency for both Spanish and English. Consistent with the graph, evidence for both the main effect of consistency and the interaction between tDist and consistency ranged from inconclusive to strong evidence in favor of the null hypothesis (*BF* = 0.05–1.21). These results indicate that consistency does not appear to enhance speech perception flexibility (see Tables A3.23–A3.26 for complete statistics).

Figure 3.22
(a,b) Recovery Rate and (c,d) Recovery Latency as a Function of tDist and Consistency for Spanish (Left Column) and English (Right Column)



Note. Lines represent model-predicted values. "Low consistency" represents one standard deviation below the mean, and "high consistency" represents one standard deviation above the mean. These splits are provided for illustrative purposes only, as consistency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

To sum up, speech perception consistency was associated with higher VOT-compatible initial lexical activation in most conditions. This effect was observed in both languages and both phoneme categories, with the exception of English /b/, which is likely more challenging due to its VOT overlap with Spanish /p/. In contrast, consistency did not facilitate speech perception flexibility as measured by recovery rate and recovery latency.

3.4 Discussion

We investigated whether speech perception gradiency and consistency influence the initial lexical activation and/or speech perception flexibility in both L1 and L2. Four key findings emerged: (1) listeners relied more on top-down lexical information in L2 than in L1, particularly those with lower proficiency; (2) consistent with Kapnoula et al.'s (2021) findings for L1 English, we found that in L1 Spanish, more gradient listeners exhibited greater speech perception flexibility when processing misleading auditory input; (3) this effect did not extend to L2 speech perception; and (4) more consistent listeners exhibited greater initial lexical activation in both L1 and L2. We also examined cross-linguistic relationships in gradiency and consistency between L1 and L2. L1 and L2 gradiency were positively correlated only among listeners with high L2 proficiency, whereas L1 and L2 consistency were positively correlated across participants regardless of proficiency.

Taken together, these findings suggest that the role of gradiency in speech perception is modulated by listeners' reliance on bottom-up versus top-down cues. Because less proficient L2 listeners rely less on bottom-up cues, likely due to less robust category representations, gradiency, which depends on such cues, becomes less impactful. In contrast, consistency appears to play a more important role in early lexical activation. Moreover, although listeners process speech differently in their two languages, L1 and L2 gradiency become increasingly aligned as L2 proficiency improves, while L1 and L2 consistency remain generally aligned.

3.4.1 Differential Effects of Gradiency Shaping L1 and L2 Speech Perception Flexibility

Participants completed the VAS and VWP tasks in both their L1 Spanish and L2 English. Our L1 results showed that, after being initially misled by garden-path stimuli, listeners with higher gradiency recovered (i.e., corrected their interpretation) more quickly than less gradient ones, especially when the acoustic deviation from the target was large. This observation aligns with Kapnoula et al.'s (2021) explanation that more gradient listeners are

better able to consider multiple hypotheses concurrently. Importantly, we do not mean that gradient listeners are inherently more capable of evaluating multiple hypotheses, but rather that their finer-grained phonetic encoding provides richer input that supports the activation of multiple lexical candidates. In this way, higher gradiency helps listeners avoid early commitment to a single phonological category and, in turn, to a single lexical interpretation, enabling more efficient resolution of misunderstandings. Our results demonstrate that the ability of more gradient listeners to recover from ambiguous or misleading information applies cross-linguistically, as we successfully extended Kapnoula et al.'s (2021) findings from English to Spanish, despite the substantial differences in VOT for stop consonants in the two languages. This reinforces the conclusion that a fundamental benefit of gradiency is its enhancement of speech perception flexibility.

In contrast to L1 Spanish, the facilitatory effect of gradiency was not observed in L2 English. The contrasting findings between L1 and L2 are intriguing and may be explained by the greater contribution of top-down (lexical) processing in L2, which is a natural consequence of less robust phonological categories and weaker gradiency in L2. Our data provide at least two pieces of evidence supporting this interpretation. First, when listening to L2 English words in the VWP, even at maximum tDist (e.g., *peachball*), participants accepted the target more often (Figure *3.8*b) than in the L1 condition (Figure *3.8*a). Second, participants exhibited a less gradient target-fixation pattern, reflecting reduced sensitivity to subphonemic differences, in L2 (Figure *3.10*b) compared to L1 (Figure *3.10*a). Therefore, it is reasonable that gradiency, being inherently tied to bottom-up acoustic processing, predicts speech perception flexibility more strongly in L1.

More broadly, this pattern aligns well with findings showing that when bottom-up processing is compromised, listeners rely more on top-down lexical knowledge (e.g., Derawi et al., 2022; Gianakas & Winn, 2019; Giovannone & Theodore, 2021a, 2021b; Mattys et al.,

2009, 2010, 2013; Mattys & Scharenborg, 2014; Mattys & Wiget, 2011; Schwartz et al., 2013), provided that L2 lexical representations are sufficiently robust (Ishida & Arai, 2016; Mattys et al., 2010; Samuel & Frost, 2015). Several of these studies suggest that reliance on lexical information can serve as a compensatory mechanism for difficulties in distinguishing acoustic information, thereby enhancing speech perception (e.g., Derawi et al., 2022; Giovannone & Theodore, 2021a, 2023). On this account, the more bottom-up processing is compromised, the more listeners are expected to rely on top-down lexical knowledge.

We argue that greater reliance on top-down information in L2 is a natural consequence of perceptual difficulties associated with processing L2 phonological contrasts. These difficulties are both theoretically and empirically well established. Theoretically, the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) posits that L2 sounds that are phonetically close to an L1 category are particularly difficult to acquire. Accordingly, English /b/, which is closer in VOT to Spanish /p/, is predicted to be more difficult for Spanish speakers to learn than English /p/. Our VAS and VWP results support this asymmetry, and empirical work has shown that Spanish–English bilinguals often struggle to maintain two distinct VOT representations for bilabial stops in English and Spanish (Williams, 1977a), though individual differences exist. Thus, Spanish–English bilinguals may not have fully differentiated phonetic categories for English /b/ and /p/. This lack of sharp category separation, as also reflected in our data, could explain why L2 listeners rely more on top-down information and, consequently, why gradiency does not facilitate speech perception flexibility in L2 to the same extent as in L1.

Despite the challenges our participants faced in processing the English bilabial stop contrast, we believe that the slope extracted from the L2 English VAS is a valid index of gradiency for several reasons. First, previous research has shown that even though L1 Spanish speakers do not produce English voiceless stops with fully L1-like VOTs, they

nonetheless develop distinct phonetic categories for the aspirated English stops (Flege & Eefting, 1987). Specifically, although participants produced /p/, /t/, /k/ in English with VOTs shorter than those of monolingual English speakers, they produced longer VOTs in English than in Spanish. These findings suggest that, despite some difficulty in fully acquiring L2 contrasts, Spanish–English bilinguals are capable of developing differentiated phonetic categories for English stop consonants. Furthermore, previous studies have demonstrated that in L2 acquisition, phonetic contrasts are easier to learn when learners only need to shift an existing L1 contrast than when they must establish an entirely new one (e.g., Chládková et al., 2022; Escudero & Boersma, 2004; Goudbeek et al., 2008). Because both Spanish and English use VOT to distinguish bilabial voicing, acquisition in this case involves shifting the category boundary along a familiar dimension rather than creating a contrast based on an unfamiliar cue—a scenario that prior work suggests is easier for L2 learners. The English /b/–/p/ contrast thus represents a case of intermediate difficulty for Spanish listeners: harder than contrasts aligned across both languages, but easier than those requiring the acquisition of novel cues. Consistent with this interpretation, and as shown in our VAS analyses, our participants are likely to have VOT representations for English /b/ and /p/ that are less robust than in L1 but sufficiently differentiated to allow reliable estimation of gradiency.

Based on our VAS analyses, we found that our Spanish-speaking participants exhibited less robust category representations for the English bilabial contrast than for the Spanish contrast, and this effect was not modulated by proficiency. Although phonetic representations might be expected to improve with increasing proficiency, we did not observe such a pattern. One likely explanation is that our participants primarily learned English in an L2-speaking environment, where they frequently encountered fellow learners and teachers producing difficult sounds in noncanonical ways (Eger & Reinisch, 2019). Developing robust

phonetic representations in L2 may also be particularly challenging without extensive immersion in an L1-speaking environment.

3.4.2 Consistency Facilitates Initial Lexical Activation in Both L1 and L2, but Not Speech Perception Flexibility

Previous work suggests that higher consistency is associated with more effective learning of unfamiliar phonetic contrasts (Fuhrmeister et al., 2023; Honda et al., 2024a) and with better reading and language abilities (H. Kim et al., 2025b–d). The present study highlights another potential advantage of consistency: its role in facilitating initial lexical activation. Specifically, we found that more consistent listeners fixated on VOT-compatible pictures more frequently before the POD. Importantly, this effect was observed in three out of four tested cases: for both phonemes in L1 Spanish and for one phoneme (i.e., /p/) in L2 English. These findings provide evidence for the role of consistency in early spoken word recognition.

Consistent with the rationale outlined in our hypothesis, we interpret this effect as follows: the speech perception system of listeners with lower consistency is likely characterized by greater noise and/or lower stability in cue-to-category mappings, resulting in increased uncertainty. Such uncertainty may lead listeners to adopt a "wait-and-see" strategy during the initial stages of spoken word recognition; that is, to delay lexical activation to avoid premature commitment until more information becomes available (McMurray et al., 2022). This interpretation is consistent with previous findings showing that listeners delay lexical activation when the input is degraded or noisy, for example in cochlear implant users and in normal-hearing listeners presented with noise-vocoded speech (Farris-Trimble et al., 2014; McMurray et al., 2017; Smith & McMurray, 2022). This "wait-and-see" strategy may therefore reflect a natural consequence of reduced stability in cue-to-category mappings or a

coping mechanism adopted by listeners under high uncertainty, enabling them to maintain alternative lexical candidates (see McMurray et al., 2022, for discussion).

Alternatively, listeners with lower consistency may activate lexical candidates in a less precise or more diffuse manner, resulting in weaker or noisier activation and more sustained competition among candidates. From this perspective, consistency may modulate the quality or selectivity of early lexical activation rather than its timing per se. The present data cannot fully distinguish between delayed activation and reduced selectivity, particularly given that the analyses focus on the early portion of the word before the point of disambiguation. Future studies will be needed to disentangle these possibilities.

Both interpretations are consistent with the idea that lower consistency reflects reduced stability in cue-to-category mappings, which introduces greater uncertainty during speech perception. This perspective aligns with the ideal observer model, which posits that high uncertainty prompts listeners to adjust their mapping of cue values onto phoneme categories, thereby allowing them to keep multiple options open until further clarifying information becomes available (Clayards et al., 2008; Nixon et al., 2016). Regardless of the specific underlying mechanism, these findings, together with prior research, suggest that consistency confers multiple benefits for speech perception.

In fact, the current findings converge with recent results reported by H. Kim et al. (2025f), who asked English-speaking participants to complete VAS tasks with different continua including stop consonants (e.g., *beach–peach*, *dime–time*) and a VWP task with minimal pairs involving the same phonetic contrast (e.g., *bin–pin*). They found that listeners with higher VAS consistency exhibited faster spoken-word recognition for stop consonants (but not for fricatives). Although the underlying mechanism remains unclear, our proposed account may offer some insight into this pattern.

While we observed a positive effect of gradiency on speech perception flexibility in L1 but not in L2, we found a largely language-general effect of consistency on the initial lexical activation across both languages. However, this effect was inconclusive for English /b/. This inconclusive pattern may reflect the phonetic proximity of English /b/ to Spanish /p/ in VOT, which could make this contrast particularly difficult for Spanish listeners to acquire. Such an account is consistent with the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021), which predicts that L2 speech sounds that are phonetically similar to an existing L1 category are especially challenging to learn. Accordingly, the absence of a consistency effect for English /b/ may reflect the greater difficulty associated with this sound, such that any facilitative role of consistency has not yet fully emerged.

3.4.3 Cross-Language Relationships in Gradiency and Consistency

Our final research question examined whether L1 gradiency and consistency are correlated with their L2 counterparts. Regarding gradiency, while the full dataset provided little evidence for a relationship (Figure *3.15*a), findings from participants with higher L2 proficiency revealed a positive correlation (Figure *3.15*b). One possible explanation for this pattern is that gradiency may influence speech processing in both L1 and L2, but its role is modulated by differences in cue reliance discussed above: higher-proficiency listeners may rely more on acoustic-phonetic cues, whereas lower-proficiency listeners may rely more heavily on lexical information. As a result, because the VAS presents consonant–vowel stimuli with limited lexical content, its sensitivity as a measure may be reduced for lower-proficiency listeners, who may be less attuned to fine-grained phonetic differences. This interpretation is consistent with Desmeules-Trudel and Zamuner (2023), who showed that L1 English–L2 French bilinguals with earlier age of exposure to French exhibited more gradient sensitivity to vowel nasalization in French than those with later exposure. It also aligns with

Honda et al.'s (2024a) proposal that differences in processing strategies between L1 and L2 may account for the weak relationship between L1 gradiency and unfamiliar sound learning.

It is important to note that we used the same stop consonant contrast across languages, which may have facilitated the emergence of cross-language gradiency alignment. If different contrasts were used, even within the same language, such alignment would be less likely. Previous studies have shown that correlations between gradiency measures are stronger when the contrasts rely on similar acoustic cues (e.g., English /b/–/p/ and English /d/–/t/, $r(64) = .41$, $p = .001$; Kapnoula & McMurray, 2021) than when they rely on different cues (e.g., English /b/–/p/ and English /s/–/ʃ/, $r(57) = .19$, $p = .16$; Kapnoula et al., 2021). Overall, these findings suggest that gradiency depends on how listeners process specific acoustic cues, supporting the view that individual differences in gradiency arise from differences in the early encoding of those cues (Kapnoula & McMurray, 2021). Nevertheless, other studies have reported higher correlations between gradiency measures derived from different languages (e.g., English /d/–/t/ and Korean /t/–/tʰ/, $R^2 = .309$; Kong & Kang, 2023) or from different contrasts within a language (e.g., English /d/–/t/ and English /s/–/ʃ/, $r(55) = .32$, $p = .02$; Fuhrmeister & Myers, 2021; English /u/–/a/ and English /d/–/g/, $r(18) = .70$; Bidelman et al., 2026).

Recall that we found a positive correlation between L1 and L2 consistency, regardless of proficiency. This finding suggests that consistency reflects an individual trait, aligning with previous research showing positive correlations in consistency across tasks (e.g., Honda et al., 2024a), phonetic categories (e.g., Fuhrmeister et al., 2023; H. Kim et al., 2025e), and modalities (e.g., visual vs. auditory; H. Kim et al., 2025d). Importantly, by including five vowels, two stops, and one fricative, H. Kim et al. (2025e) reported that the average correlation for consistency ($r = .67$) was nearly twice that for slope ($r = .36$). The current study extends these findings by demonstrating cross-language consistency in speech

perception when processing the same acoustic cue but in different languages. This suggests that consistency may represent a more general mechanism than gradiency.

3.4.4 Conclusion

The current study provides evidence that both gradiency and consistency play important roles in spoken word recognition. Consistency appears crucial during the early stages of spoken word recognition: listeners with lower consistency may delay lexical activation or activate lexical candidates less selectively. Notably, this effect emerged in both L1 and L2. Conversely, gradiency enhances speech perception flexibility by facilitating recovery from misleading auditory input, and this effect was observed only in the L1. This pattern is likely driven by differences in listeners' reliance on bottom-up versus top-down information; when processing L2 speech, listeners depend more strongly on top-down lexical information, rendering gradiency less relevant. Collectively, these findings elucidate the distinct functional roles of gradiency and consistency in speech processing across L1 and L2 and highlight the importance of proficiency in strengthening bottom-up cue use in the L2.

4 Functions of Consistency and Gradiency: Perception–Production Alignment

Parts of this chapter are based on:

Wong, B. W. L., Samuel, A. G., & Kapnoula, E. C. (under review). *Weak links between speech perception and speech production at the subphonemic level. Preprint:* https://doi.org/10.31234/osf.io/qr9zt_v1

4.1 Introduction

This final experimental chapter examines another potential function of speech perception gradiency and consistency: speech sound alignment. Specifically, we explore the relationship between speech perception and speech production, asking whether listeners who exhibit greater gradiency in perception imitate speech sounds more accurately and/or in a more gradient manner in production. We also examine whether listeners with higher speech perception consistency tend to produce speech sounds more consistently.

Speech alignment, the tendency for interlocutors to adjust their speech to one another, is a fundamental aspect of spoken communication and social interaction (Giles et al., 1991). A common form of alignment is speech imitation (sometimes referred to as convergence), which is relatively stable within interlocutors (Wade et al., 2021) but varies in degree across them (Pardo et al., 2018). Phonetic imitation, as demonstrated by Goldinger (1998) through episodic memory effects in speech shadowing, provides evidence that perception can directly influence production.

Several theoretical frameworks, including the interactive alignment framework and simulation theory, propose that speech imitation arises from automatic priming of shared representations across multiple linguistic levels, from situation models to syntax, semantics, and phonology (Gambi & Pickering, 2013; Garrod & Pickering, 2004; Pickering & Garrod, 2004). Although these frameworks do not explicitly address individual differences in subphonemic sensitivity, they imply that the ability to perceive subtle speech contrasts is reflected in production. In principle, individuals with more finely tuned acoustic-phonetic (e.g., more gradient) representations are better equipped to detect these distinctions in speech and reproduce them accurately.

Models of speech motor control have proposed similar ideas (Hickok et al., 2011; Houde & Nagarajan, 2011; Tourville & Guenther, 2011). These frameworks posit that speech

perception plays a key role in production through an auditory feedback mechanism that informs motor control. Speech production is guided by perceptual targets; the goal of production is to generate a specific sound sequence that matches stored auditory representations. These representations are thought to be initially acquired through speech perception, suggesting that individual differences in perception may influence production targets. Over time, speech production may be fine-tuned through auditory feedback: when the system detects deviations between the produced sound and the target representation, compensatory adjustments are made. Following this rationale, individuals with greater gradiency and/or consistency may be more sensitive to such discrepancies, which could drive the speech production system to become more precise and less variable, thereby yielding greater consistency in imitation.

Perkell (2012) proposed that speech production targets are represented, at least in part, in acoustic terms. Under this account, individual differences in auditory perception can influence the variability in the precision of speech production targets. Listeners with greater auditory discrimination are able to encode finer acoustic distinctions, which results in more sharply specified perceptual targets. These more precise targets provide clearer goals for the motor system, leading to less variable speech production. This view is supported by empirical work on vowels (e.g., Franken et al., 2017; Perkell et al., 2004, 2008), which demonstrated that individuals with better speech-sound discrimination exhibited smaller within-phoneme variability (i.e., less dispersion) and greater between-phoneme separation in production, as assessed in pseudoword reading tasks. While these studies demonstrate that individual differences in auditory discrimination are reflected in the precision of speech production targets, they primarily address variability between phoneme categories (e.g., /i/ vs. /ɪ/). Much less is known about whether similar relationships exist within phoneme categories at the subphonemic level, such as sensitivity to fine-grained VOT differences within /b/ or /p/.

Despite the theoretical links proposed in the studies above, empirical findings on the relationship between speech perception and speech production, particularly for VOT, have been mixed. For instance, Bailey and Haggard (1973) used a nine-point rating scale to assess the perception of /k/–/g/ continua and asked participants to produce sentences such as *"It's a g/kill."* Across a range of perceptual and production measures, they found no statistically significant correlations. Similarly, Samuel (1979) did not find a relationship between the perceived prototype (i.e., the stimulus listeners judged as the most representative exemplar of the category) and the production of the /ga–ka/ contrast. However, subsequent work by Newman (2003) examined correlations between perceptual prototypes for stop consonants (e.g., /pa/) and their average productions, and found that listeners whose perceptual prototypes had longer VOTs also tended to produce longer VOTs. This finding provides clearer evidence for a perception–production correspondence. In addition, Nielsen (2011) showed that English speakers produced significantly longer VOTs after exposure to speech with extended VOTs but did not shorten their productions after exposure to reduced VOTs, suggesting that phonetic imitation is selective rather than fully automatic (see also Lindsay et al., 2022). Importantly, none of these studies included explicit imitation instructions. Without such instructions, imitation may reflect a listener's propensity to imitate, shaped by social or personality factors (e.g., openness to experience; see below), rather than their ability to imitate. This likely represents one reason for the mixed findings observed in the literature. Supporting this interpretation, Schertz and Paquette-Smith (2023) used an explicitly instructed imitation task and found phonetic convergence to both lengthened and shortened VOTs, indicating that when imitation is intentional, convergence occurs bidirectionally rather than selectively.

Mixed findings have also been reported in studies focusing on acoustic-phonetic (subphonemic) variation in vowels. H. S. Cheng et al. (2021) asked participants to repeatedly

produce the words *head* and *had* and then complete a 2AFC identification task using a *head–had* continuum. The researchers assessed perception using a measure of categorical labeling consistency and production using measures such as vowel contrast distance, vowel space area, and centering ratio (i.e., the degree of change in variability over time). They found no statistically significant correlations between perception and production measures. However, they emphasized that a perceptual task tapping within-category discrimination ability, rather than categorical identification tasks such as 2AFC, might be better suited to capture perception–production links. Task sensitivity, therefore, likely represents a second factor contributing to the mixed findings in the literature. The present study includes the VAS task, which provides a more sensitive index of gradiency and consistency in speech perception, to address this limitation.

Although previous studies have manipulated subphonemic cues, they typically examined perception and production using categorical measures. To our knowledge, few studies have directly tested the relationship between perception and production using continuous, within-category measures. One relevant exception is a study by D. Kim and Clayards (2019), who investigated vowel imitation using both spectral and durational cues that varied continuously within categories. Participants completed a 2AFC perception task in which cue weighting was modeled as a continuous measure of perceptual acuity, and an imitation task in which participants reproduced vowels differing across small spectral steps. The researchers found that individuals with greater perceptual acuity (indexed by stronger spectral cue weighting) also demonstrated better imitation of vowel duration. Although their perceptual and production measures tapped different acoustic dimensions, this finding nevertheless suggests a cross-modal link between speech perception and production mechanisms at the subphonemic level.

Besides speech perception abilities, other factors have also been shown to affect speech imitation. For example, openness to experience (one of the Big Five personality traits) and attention-switching ability were found to predict implicit imitation of words (Yu et al., 2013). Listeners with higher openness to experience and better attention-switching skills exhibited greater imitation. Another study found that musical perception skills, articulatory ability, openness to experience, and self-rated imitation difficulty predicted accent imitation during a story-reading task, as rated by external listeners (Myers et al., 2024a). Specifically, listeners with better musical perception skills, better articulatory ability, higher openness to experience, and those who perceived the imitation task as less difficult tended to imitate accents to a greater degree. However, these studies examined imitation at a holistic level (e.g., word-level or accent-level imitation) rather than at the subphonemic level, so it is possible that these factors are not directly relevant to the present study. Nonetheless, we still measured them as control variables.

Notably, Myers et al. (2024a) also examined the relationship between VAS measures (i.e., gradiency and consistency) and accent imitation but found no statistically significant correlation. This may be because the VAS primarily indexes sensitivity to phoneme-level distinctions, whereas accent imitation involves broader speech patterns. This difference in the stimuli used across tasks likely represents a third factor contributing to the weak or inconsistent perception–production links reported in the literature. In contrast, the present study focuses on the explicit (i.e., deliberate) imitation of consonant-vowel syllables rather than accent imitation, thereby aligning the production stimuli with those used in the VAS task.

As discussed throughout this dissertation, listeners are generally gradient, though individual differences exist. Previous research has also shown that listeners can maintain this gradiency for a relatively long period, at least one second (Brown-Schmidt & Toscano, 2017;

Falandays et al., 2020; Sarrett et al., 2020). Beyond perception, findings are mixed regarding whether L1 Spanish speakers can imitate subtle VOT differences within their L1 phonological categories. Olmstead et al. (2013) showed that speakers can exhibit gradient subphonemic production, producing a continuous increase in VOT as stimulus VOT increases, even if the exact values are not matched. In contrast, Flege and Eefting (1988) reported a more categorical imitation pattern clustered around prototypical /b/ and /p/ values. These discrepancies likely reflect individual differences, but this possibility has not been systematically examined. Based on the frameworks outlined above (Hickok et al., 2011; Houde & Nagarajan, 2011; Perkell, 2012; Tourville & Guenther, 2011), imitation gradiency and imitation consistency may vary across individuals, and such variability is possibly linked to individual differences in the corresponding constructs within the speech perception system. By investigating individual differences in gradiency and consistency on both the perception and production sides, the present study aims to extend existing theoretical frameworks to determine whether speech imitation can operate at the subphonemic level.

To address the three potential reasons for the weak links observed between speech perception and speech production, specifically (1) social or personality factors that influence one's willingness rather than ability to imitate, (2) the use of insensitive tasks that fail to capture within-category sensitivity, and (3) the use of different stimuli across perception and production tasks, the present study adopts a more controlled design. Unlike most previous work, the current study (1) includes an explicit (deliberate) imitation task to capture imitation ability rather than propensity (D. Kim & Clayards, 2019), while still accounting for individual factors related to the latter, such as articulatory ability, musical perception skill, openness to experience, attention switching, and perceived imitation difficulty; (2) employs a VAS task, which reliably measures gradiency and consistency (Honda et al., 2024a; H. Kim et al., 2025e; Kong & Edwards, 2016); (3) uses identical stimuli, focusing on subphonemic

differences in Spanish stop consonants (/b/–/p/), across the perception and imitation tasks. In addition, a temporal order judgment task assessing auditory acuity was included due to its established link to gradiency (Kapnoula & Samuel, 2024). A visual working memory task (nonlinguistic and nonauditory in nature) was also included to test whether the relationship between consistency in speech perception and production, if found, is speech-specific rather than driven by general differences in effort between participants. With these controls in place, this study adopted an individual-differences approach to examine whether gradiency and consistency explain additional variance in imitation accuracy, gradiency, and consistency beyond the predictors related to speech imitation identified in previous research.

Specifically, we asked two research questions: (1) Do listeners with greater gradiency in speech perception imitate sounds more accurately and/or more gradiently in speech production? (2) Do listeners with higher speech perception consistency imitate sounds more consistently? To address the first question, we extracted three indices of imitation ability: (1) imitation accuracy, defined as the difference between the presented (i.e., stimulus) VOT and the imitated VOT after accounting for participants' baseline production; (2) simple imitation accuracy, which uses the same calculation but without baseline adjustment; and (3) imitation gradiency, defined as the inverse of the imitation slope across VOT steps. Although both accuracy and gradiency involve continuous stimuli, they capture distinct components of imitation. Accuracy measures how close productions are to the target VOTs. Gradiency measures how well productions track changes across the continuum. Because of the continuous design, high accuracy implies high gradiency, but the reverse is not necessarily true: a listener can vary their productions across steps (gradient imitation) yet still be inaccurate overall if their productions are consistently shifted.

Based on the interactive alignment framework and simulation theory, we hypothesized that listeners with greater speech perception gradiency would imitate more

accurately (i.e., produce sounds acoustically closer to the prompts) and/or more gradiently (i.e., showing a shallower imitation slope across VOT steps). Furthermore, drawing on models of speech motor control, we hypothesized that listeners with higher perceptual consistency[6] would produce speech sounds with greater consistency. This prediction is also supported by evidence that consistency appears to be a more stable and general individual trait than gradiency, showing positive correlations across tasks (e.g., Honda et al., 2024a), phonetic categories (e.g., Fuhrmeister et al., 2023; H. Kim et al., 2025e), and modalities (e.g., visual vs. auditory; H. Kim et al., 2025d).

4.2 Method

4.2.1 Participants

We initially planned to recruit 80 participants based on a power analysis (Appendix 1.3). Ultimately, 100 participants were recruited because 19 participants exhibited relatively narrow response ranges in the VAS task (curve-fitting $R^2 < .5$)[7], and one participant was excluded due to a technical error; these participants were replaced. The final sample therefore consisted of 80 Spanish speakers (59 females; 21 males) aged between 18 and 35 years (M = 24.4; SD = 4.6) who had acquired Spanish before age 4. All participants demonstrated high proficiency in Spanish, scoring over 92% on the picture-naming task administered during registration in the BCBL participant database (de Bruin et al., 2017). They reported no history of hearing, reading, motor, or speech disorders, and no neurological conditions or head injuries. The study was approved by the BCBL Ethics Review Board (ref. 040225SM)

[6] We use the term "perceptual consistency" to draw a direct contrast with "production consistency," referring to the stability of responses in the perception task rather than to consistency at the level of perceptual encoding.
[7] The high number of flat curves was unexpected, given that the task involved participants' L1 and that the stimuli were produced by the same speaker and employed the same manipulation as in other experiments presented in this dissertation. The most plausible explanation is the use of a gender-neutral voice (as described below), which may have been less familiar to participants and consequently made the VOT contrasts more difficult to distinguish.

and adhered to the principles of the Declaration of Helsinki. All participants provided written informed consent and received monetary compensation for their participation.

4.2.2 Task Overview

Participants completed a VAS task, an imitation task, a temporal order judgment task, an articulatory task, a musical perception task, a visual working memory task, and a language and musical background questionnaire in a sound-attenuated booth during a one-hour behavioral session. The tasks were administered in this fixed order. All auditory stimuli were presented through Beyerdynamic DT 770 Pro (250 Ω) headphones powered by an M-Audio AIR 192|4 amplifier/receiver. Visual stimuli were displayed on a 24-inch monitor operating at a resolution of 1024 × 768 pixels. All audio recordings for the imitation and articulatory tasks were made using a Samson C01U Pro microphone. Table *4.1* summarizes all tasks and variables.

Table 4.1
Summary of Tasks and Variables for Experiment 3

Order	Task	Variable	Duration (min)
1	VAS	Speech perception gradiency and consistency	10
2	Imitation	Imitation accuracy, simple imitation accuracy, imitation gradiency, and imitation consistency	15
3	Temporal order judgment	Auditory acuity	4
4	DDK	Articulatory ability	1
5	Corsi block-tapping	Visual working memory	5
6	Mini-PROMS	Musical perception skills	15
7	Questionnaire	Openness to experience, attention switching, imitation difficulty	5

Note. VAS = Visual Analog Scale; DDK = Diadochokinetic rate.

4.2.3 VAS

The VAS stimuli consisted of seven sounds forming a /ba/–/pa/ continuum, with VOTs ranging from −28 ms to 17 ms (−28, −20, −13, −5, 2, 9, and 17 ms). The stimulus manipulation and experimental procedures were similar to those described in Section 3.2.3 except for four key differences.

First, to prevent participants from imitating the speaker's gender-specific characteristics, the female voice recordings were transformed into a gender-neutral voice using the Change Gender function in Praat (formant shift ratio = 0.88; new pitch median = 150 Hz). Because the same stimuli were used for both the VAS and imitation tasks, the VAS stimuli were also processed using the same transformation. A pilot test confirmed that the resulting voice sounded sufficiently gender-neutral.

Second, all stimuli had identical vowel durations. Vowel duration was not manipulated because, unlike in English, where vowels following voiced consonants tend to be longer than those following voiceless consonants (Allen & Miller, 1999), this relationship in Spanish is weaker and less consistent (Hammond, 2001; Hualde, 2005; Zimmerman & Sapon, 1958). Third, F_0 was held constant at 186 Hz and was not manipulated. These design choices simplified the paradigm and isolated the effects of VOT, ensuring that any observed differences could be attributed specifically to this cue.

Finally, the assignment of response labels was counterbalanced within participants to control for potential response biases. Each participant completed both label mappings: two blocks with "ba" on the left and "pa" on the right, and two blocks with the reverse arrangement. Half of the participants began with the "ba-left" condition, and the other half with the "pa-left" condition. The task comprised 140 trials in total, with 35 trials per block.

Five practice trials were included before each label mapping. The task took around 10 minutes.

4.2.4 Imitation Task

The stimuli used in the imitation task were identical to those used in the VAS task. Participants first produced the syllables /ba/ and /pa/ naturally eight times each toward the microphone, serving as a baseline measure, while the corresponding visual labels ("ba," "pa") and a *siguiente* ("next") button below the label were displayed on the screen. They clicked *siguiente* once they had finished each recording. In the main task, participants listened to each stimulus twice, separated by a 700-ms interstimulus interval (ISI). During each listening phase, an ear symbol appeared on the screen and remained visible for 1,430 ms. Then, a red circle appeared on the screen, accompanied by the *siguiente* button below it, prompting participants to imitate the sound once. Participants were instructed to click *siguiente* only after completing their production. To prevent premature responses, the *siguiente* button became clickable only after two seconds had elapsed in each trial. The number of stimulus presentations matched that in the VAS task (20 repetitions per stimulus). The main task comprised four blocks of 35 trials, totaling 140 trials, and lasted approximately 15 minutes. Participants' productions were recorded. The task was implemented in jsPsych (version 7.3.1; de Leeuw et al., 2023) and administered via the Chrome browser.

4.2.5 Temporal Order Judgment Task

This task, adapted from Kapnoula and Samuel (2024) and following the procedure of Simon and Winkler (2018), measured auditory acuity in terms of temporal order judgment. Temporal order judgment was chosen because it reflects a timing-based skill closely related to VOT perception. Auditory acuity has been found to correlate with gradiency in Spanish speakers (Kapnoula & Samuel, 2024), but its relationship with consistency remains unknown.

In this task, pairs of 10-ms, 800-Hz pure tones were presented to participants. The two tones in each pair were separated by a variable interstimulus interval (ISI) ranging from 5 to 150 ms, with each tone played in a different ear. Participants responded by pressing the left arrow key if the first tone was heard in the left ear and the right arrow key if it was heard in the right ear. Each participant's temporal threshold (i.e., ISI) was determined using a three-down, one-up staircase procedure. The main task began with an ISI of 100 ms. The ISI decreased by a step size after three consecutive correct responses and increased by the same step size after an incorrect response. The step size began at 20 ms and, at each downward adjustment, was reduced by 50% until reaching a minimum of 5 ms. For example, the ISI would decrease from 100 ms to 80 ms (step size = 20 ms), then later to 70 ms (step size = 10 ms), and then to 65 ms (step size = 5 ms) on the trials corresponding to each of these downward steps. The intertrial interval was between 600 and 900 ms (jittered). The task ended once the participant made eight incorrect responses in total. The auditory acuity threshold was calculated as the average ISI across the last six incorrect trials, with lower thresholds indicating higher auditory acuity. For interpretability, the values were multiplied by −1 so that higher scores reflect higher auditory acuity. This task took approximately four minutes.

Before the main task, participants completed 18 practice trials with feedback. ISI decreased across the practice phase: 150 ms for the first six trials, 125 ms for the next six, and 100 ms for the final six.

Experiment Builder (version 2.4.1; SR Research Ltd.) was used to program and run the VAS and the temporal order judgment task.

4.2.6 Diadochokinetic Rate (DDK) Task

The DDK task assessed participants' maximum speech rate as an indicator of articulatory ability. This task is commonly used to evaluate speech production disorders

(Kent et al., 2022; Maas, 2017). Participants were instructed to repeat single syllables (/pa/, /ta/, /ka/) and a trisyllabic sequence (/pataka/) into the microphone as quickly as possible for five seconds. The corresponding labels were displayed on the screen, with a *siguiente* ("next") button below. The experimenter remained in the testing room, instructed participants when to start and stop, timed each trial, and clicked *siguiente* once the recording was completed. If a participant laughed, hesitated, or ended the sequence prematurely, the trial was immediately repeated. Participants' productions were recorded. The task lasted approximately one minute. It was implemented in jsPsych (version 7.3.1; de Leeuw et al., 2023) and administered via the Chrome browser.

4.2.7 Corsi Block-Tapping Task

The Corsi block-tapping task was identical to that described in Section 3.2.7, except that it was implemented in OpenSesame version 4.0.

4.2.8 Mini-PROMS Test

The Mini-PROMS test is a reliable and validated measure of musical perceptual skills on a gradient scale, administered in an online format (Zentner & Strauss, 2017). It comprises four subtests assessing melody (discrimination of similar melodic passages; 10 trials), tuning (the ability to identify whether a note is in tune or out of tune within a chord; 8 trials), beat (discrimination of rhythmic patterns; 10 trials), and tempo (judging differences in the overall speed of musical passages; 8 trials). In each subtest, participants completed discrimination trials in which they listened to two short musical recordings and indicated their decision and confidence on a five-point scale: definitely different, probably different, I don't know, probably the same, or definitely the same. The entire task comprised 36 trials and took approximately 15 minutes to complete.

4.2.9 Questionnaire

The questionnaire included the Openness to Experience subscale of the Big Five Inventory (John et al., 1991) and the Attention Switching subscale of the Autism-Spectrum Quotient (Baron-Cohen et al., 2001). Participants also rated the difficulty of the imitation task. Additionally, the questionnaire collected information about participants' linguistic profile, including their listening proficiency in and exposure to Spanish, Basque, and English, and their musical background (i.e., duration of musical training). It took approximately five minutes to complete. Participants also rated the difficulty of the imitation task.

4.2.10 Extracting Speech Perception Gradiency and Consistency

All procedures were identical to those described in Section 3.2.10. The only differences were that the "pa-left" VAS responses were reversed so that higher values corresponded to higher /p/ ratings prior to curve fitting, and that a standard logistic function was used instead of a rotated logistic function because only VOT was manipulated. The model fit was good (average R^2 = .83).

4.2.11 Extracting Imitation Measures

The data obtained from the imitation task were the audio recordings stored as Base64 strings in an Excel file. These were automatically converted into WAV files using an R script. Each audio file was then analyzed using Dr. VOT, a software package for automatic VOT measurement (Shrem et al., 2019). Manual adjustments were made using a Praat script when necessary to ensure accuracy. The adjustment procedure followed that of a previous study (Souganidis et al., 2024). Specifically, VOT measurements were obtained via visual inspection of the spectrogram and waveform. For voiced plosives, prevoicing (i.e., negative VOT) was measured as the interval between the onset of periodic voicing and the release burst. For voiceless plosives, VOT was measured as the interval between the burst and the

onset of the following vowel. Overall, approximately 35% of trials required manual adjustment.

Some trials were excluded during manual inspection. Silent trials (0.26% of the 12,480 trials) were discarded, and trials containing unintended syllables (i.e., /da/, /ga/, /fa/, or /ta/) were excluded (0.56%). In total, including baseline trials, 12,379 observations (participant × item) remained for analysis. A small proportion of trials lacked a clear /b/ or /p/ release burst and contained only the vowel /a/. These trials were retained and assigned a VOT value of 0 ms (1.43%), as such responses likely reflected participants' attempts to imitate acoustically ambiguous stimuli rather than recording errors.

Based on these VOT measurements, four imitation measures were derived. The first measure, imitation accuracy, was initially computed as the mean absolute difference between the imitated VOT and the stimulus VOT, with larger differences indicating lower accuracy. To account for participants' baseline productions, imitation accuracy was then adjusted using the following formula: $|VOT_{target} - VOT_{baseline}| - |VOT_{target} - VOT_{imitation}|$ (Babel, 2012; Phillips & Clopper, 2011; Walker & Campbell-Kibler, 2015). Because participants produced only clear /b/ and clear /p/ sounds in the baseline phase, the baseline VOT for each continuum step was estimated using linear interpolation as follows: baseline VOT for each step = $VOT_b + ((VOT_p - VOT_b)/(6)) \times (step - 1)$, where VOT_b and VOT_p represent the average VOTs for clear /b/ and clear /p/, respectively, and the denominator (6) reflects the six equal intervals separating the seven continuum steps. A positive imitation accuracy value therefore indicates that the participant's imitation was closer to the target than their baseline production (i.e., higher accuracy), whereas a negative value indicates that the imitation deviated further from the target (i.e., lower accuracy).

The second measure, simple imitation accuracy, was defined as the average absolute difference between the imitated VOT and the stimulus VOT, without taking baseline productions into account. Because larger values indicate lower accuracy, the scores were multiplied by −1 so that higher values consistently reflect higher simple imitation accuracy for interpretability.

The third measure was imitation gradiency. Participants' imitation responses were fitted using the same logistic function as in the VAS task, and the slope of the fitted curve was extracted as a measure of gradiency. Unlike the VAS task, there was no fixed 0–100 response scale; therefore, for each participant, the minimum and maximum values for curve fitting were defined as the mean imitated VOTs for the first and last continuum steps, respectively. Twelve participants were excluded from this analysis because their fitted curves exhibited poor fit quality ($R^2 < .5$), typically due to flat response patterns. This likely reflects task difficulty rather than lack of engagement, as these participants showed well-fitted curves in the VAS task. Overall, the model fit was good ($R^2 = .67$). For interpretability, the slope values were multiplied by −1 so that higher values represent greater imitation gradiency.

Finally, imitation consistency was calculated as the standard deviation of production VOT values across tokens with the same stimulus VOT, where higher standard deviation indicates lower consistency. The values were multiplied by −1 so that higher scores reflect greater consistency.

4.2.12 Extracting Articulatory Ability

Similar to the imitation data, the four DDK recordings per participant were initially stored as Base64 strings, which were automatically converted into WAV files using an R script. These files were then analyzed using DDKtor, which generated Praat TextGrid files with each consonant and vowel annotated within each recording (Segal et al., 2022; Segal-

Feldman et al., 2025). An R script was subsequently used with the rPraat package (version 1.3.2.1; Bořil & Skarnitzl, 2016) to extract the number of consonants and vowels produced during the first five seconds, counted from the onset of the first consonant. When discrepancies occurred between the two counts (e.g., a consonant detected without its corresponding vowel), the smaller value was used. The DDK score was calculated as the average number of repetitions across the four items (/pa/, /ta/, /ka/, and /pataka/). Data from one participant were excluded because she was unable to produce speech continuously for five seconds. Two additional participants were excluded due to technical errors (data from these three participants were retained for the other tasks). For another three participants the /pa/ recording was unusable due to the same type of technical issue; in these cases, the remaining three items were averaged to obtain the score. A higher average number of repetitions reflected a higher maximum speech rate and, consequently, greater articulatory ability.

4.2.13 Statistical Analyses

We used Bayesian correlations implemented in the BayesFactor package (Morey & Rouder, 2024) to quantify evidence for both the presence and absence of relationships between variables. Correlations were computed using pairwise deletion (i.e., all participants with valid data for each variable pair were included). Bayes factors (BF) were interpreted according to the scheme described in Section 2.2.6.

Hierarchical regression analyses were conducted within a frequentist framework, as this approach allows straightforward interpretation through p-values and ΔR^2 and is more practical for models with several predictors. All dependent variables were rank-inverse-normal transformed to address mild deviations from normality and to ensure consistency across models. Diagnostic tests confirmed that all models satisfied normality and homoscedasticity assumptions. Variance Inflation Factors (VIFs) for all predictors were well

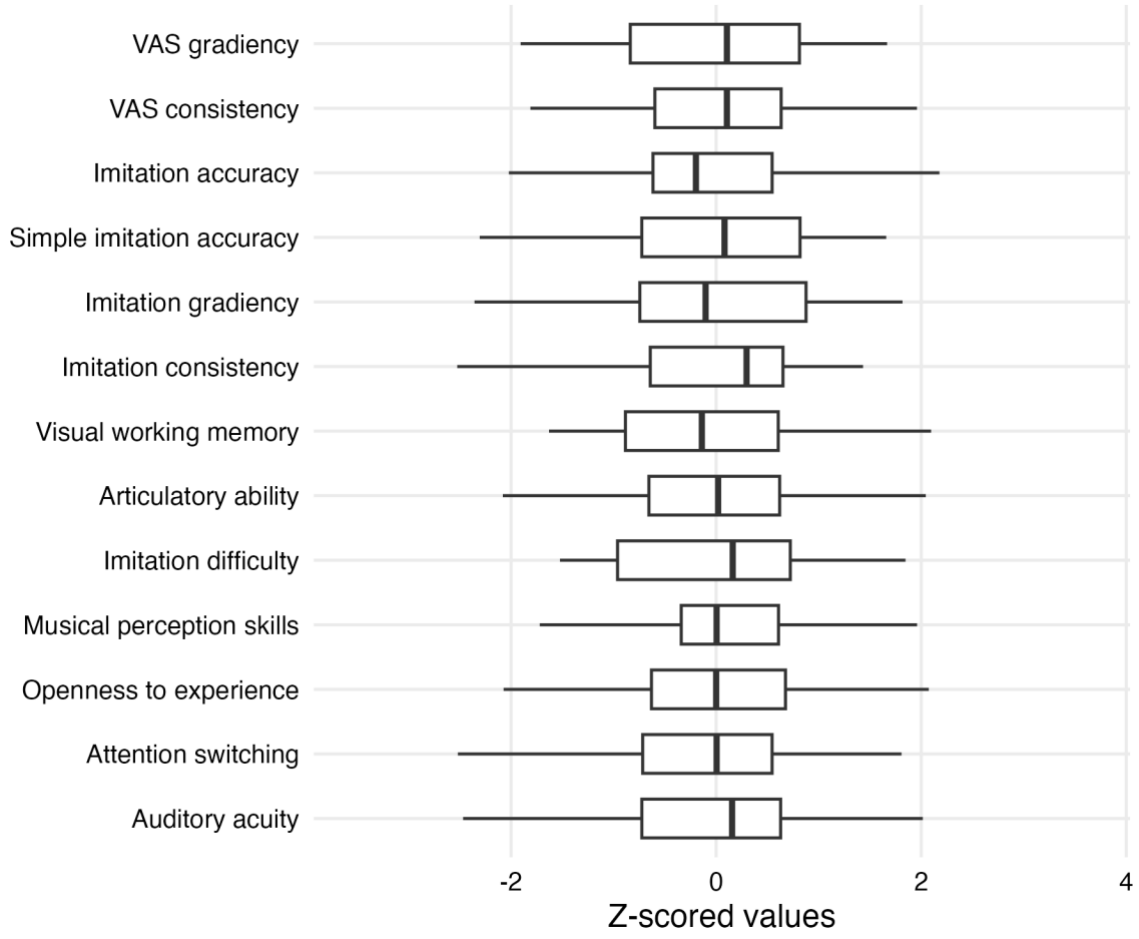
below the conventional cutoff of 5, indicating no evidence of multicollinearity. Model fit improvement was assessed by examining the change in explained variance (ΔR^2) and the significance of the added predictor in Step 2 ($p < .05$).

Four hierarchical regression models were fitted. The first three models corresponded to our first research question: Do listeners with greater gradiency in speech perception imitate sounds more accurately and/or more gradiently in speech production? In Model 1, the dependent variable was imitation accuracy. Step 1 included perceived imitation difficulty, auditory acuity, musical perception skill, articulatory ability, openness to experience, attention switching, and visual working memory as predictors. Step 2 added speech perception gradiency extracted from the VAS. Models 2 and 3 followed the same structure but used simple imitation accuracy and imitation gradiency as dependent variables, respectively. In Model 4, the dependent variable was imitation consistency, with the same predictors in Step 1 and perceptual consistency (from the VAS) added in Step 2. This final model addressed our second research question: Do listeners with higher speech perception consistency imitate sounds more consistently?

4.3 Results

Before addressing the research questions, we first examined the descriptive statistics of all variables included in the analyses. Table *4.2* reports the summary statistics based on the raw values, and Figure *4.1* visualizes the distributions using z-scored values to place all measures on a common scale for comparison. As shown, there was substantial individual variability across all measures, as desired for the correlational analyses.

Figure 4.1
Boxplots Showing the Distributions of All Variables included in the Regression Analyses



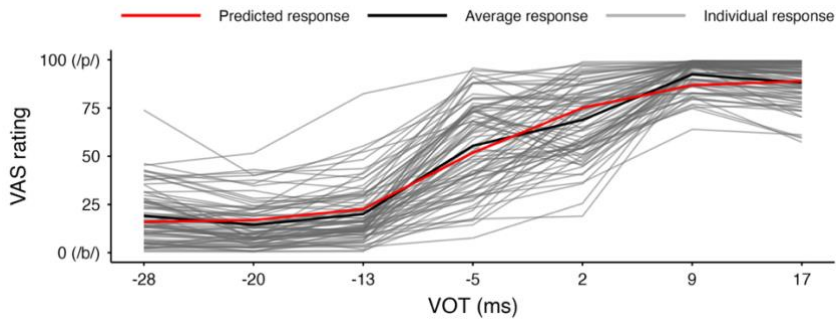
Notes. Variables were standardized (z-scored) for visualization, so that measures with different units could be displayed on a common scale. Each box represents the interquartile range (IQR), the horizontal line indicates the median, and whiskers extend to 1.5 × IQR.

Table 4.2
Descriptive Statistics for All Variables Included in the Hierarchical Regression Models

Variable	*N*	*M*	*SD*	Min	Max
VAS gradiency	80	−53.5	22.1	−95.6	−16.7
VAS consistency	80	−19.5	4.6	−35.8	−10.6
Imitation accuracy	80	0.2	19.7	−39.7	72.6
Simple imitation accuracy	80	−36.7	17.1	−76.2	−8.3
Imitation gradiency	68	−79.5	36.4	−165.3	−13.2
Imitation consistency	80	−39.6	11.2	−70.9	−23.6
Visual working memory	80	6.2	1.3	2.0	9.0
Articulatory ability	77	32.7	3.3	23.5	39.5
Imitation difficulty (1: very easy; 7: very difficult)	80	3.7	1.8	1.0	7.0
Musical perception skills	80	20.0	4.3	9.5	31.5
Openness to experience	80	36.5	5.5	25.0	48.0
Attention switching	80	32.0	5.5	18.0	42.0
Auditory acuity	80	−69.7	25.1	−131.7	−19.2

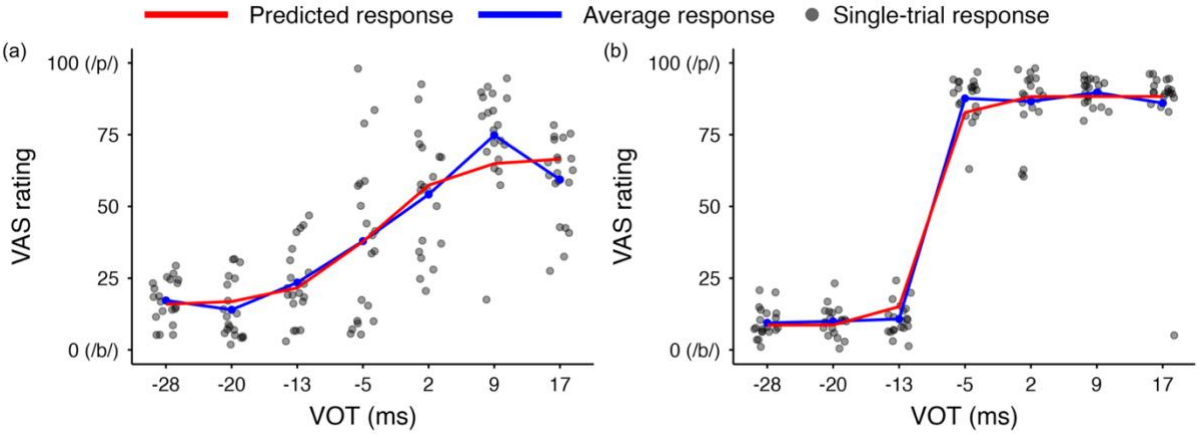
Then, we examined the VAS data. Figure *4.2* shows participants' average VAS ratings across the seven VOT values. As expected, listeners were generally gradient, exhibiting a relatively shallow slope in the mean curve; however, individual variability was also evident. As illustrated in Figure *4.3*, some participants were more gradient than others. This figure also demonstrates variability in consistency, with some participants showing higher consistency than others, with tighter clustering of dots around the fitted curve at the same VOT value indicating higher consistency.

Figure 4.2
VAS Rating across VOT



Note. The black line represents the grand average of participants' observed VAS ratings across VOT values, while the red line represents the predicted ratings based on the fitted logistic function. Grey lines show individual participants' observed curves.

Figure 4.3
Two Representative Participants Showing (a) Low Consistency and High Gradiency, and (b) High Consistency and Low Gradiency in the VAS

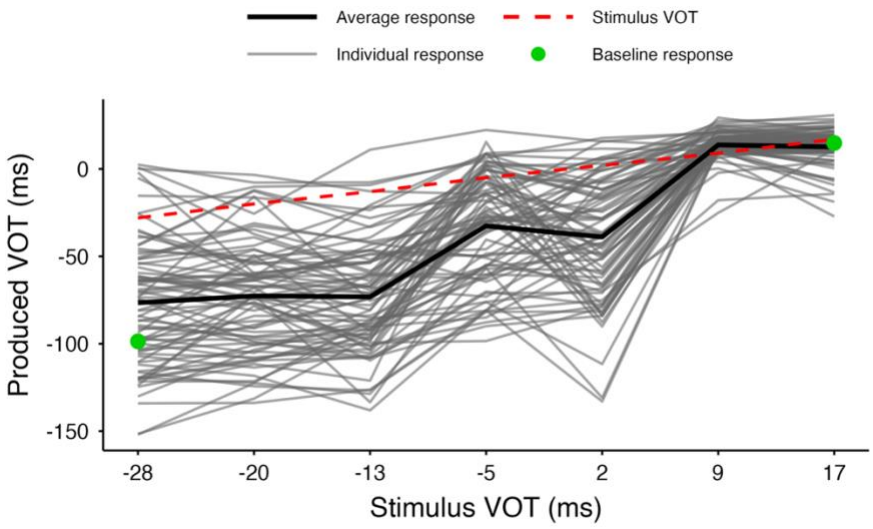


Note. The blue line represents the average of each participant's observed VAS ratings across VOT values, while the red line represents the predicted ratings from the fitted logistic function. Each dot indicates an individual trial response.

Next, we report the imitation data. Figure *4.4* shows participants' produced VOTs as a function of stimulus VOT. As indicated by the gradually increasing average curve, participants were generally gradient in their imitation patterns. They imitated stimuli with higher VOTs (i.e., on the /p/ side) quite accurately, with productions closely matching the corresponding stimulus VOTs, likely because these stimuli (e.g., extreme /p/ = 17 ms) were

acoustically close to their baseline /p/ productions (*M* = 14.8 ms, *SD* = 5.1 ms). In contrast, imitation of stimuli with lower VOTs was less accurate, as these stimuli (e.g., extreme /b/ = −28 ms) were farther from the baseline (*M* = −98.6 ms, *SD* = 30.4 ms). These baseline values align with previous VOT data in the same population (Souganidis et al., 2024). Another likely contributor to the difference between /b/ and /p/ imitation is production variability: as shown by the SD values reported above and individual curves in Figure *4.4*, /b/ productions were more variable than /p/ productions, consistent with previous findings (e.g., Lisker & Abramson, 1964). Overall, these results suggest that imitation accuracy in this task primarily reflects participants' ability to approximate the target VOTs for lower-VOT (/b/) and ambiguous stimuli, whereas imitation accuracy for higher-VOT (/p/) stimuli was near ceiling.

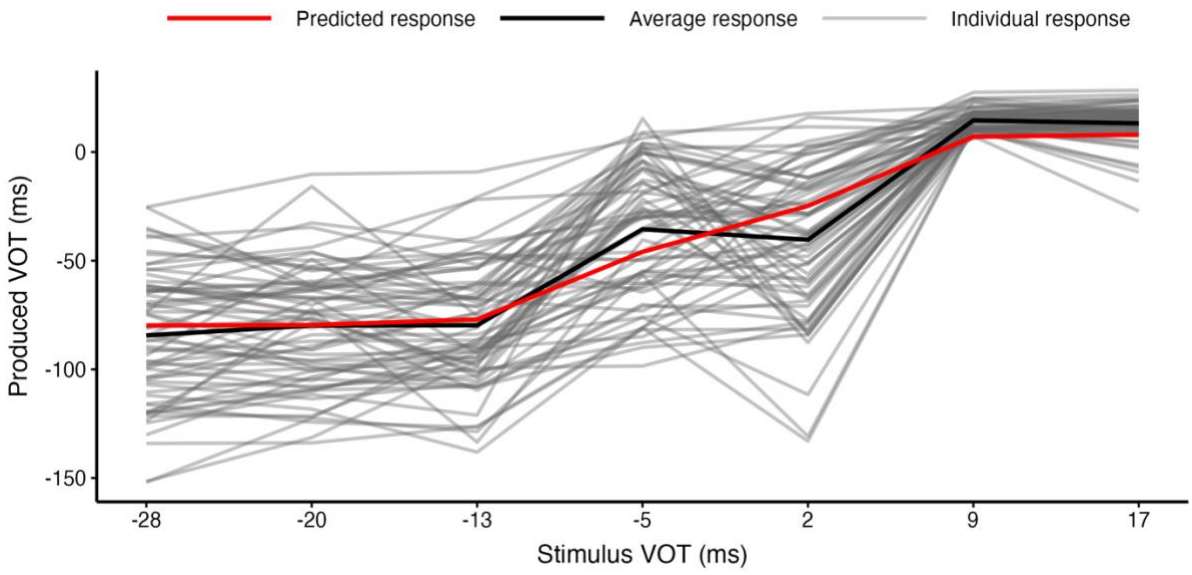
Figure 4.4
Produced VOTs as a Function of Stimulus VOT in the Imitation Task



Note. The black line represents the grand average of participants' produced VOTs across stimuli with different VOT values, while the red dashed line represents the stimulus VOTs they heard (identical to the x-axis values). Grey lines show individual participants' average produced-VOT curves. The two green dots represent the grand averages of participants' produced /ba/ and /pa/ VOT as baselines.

Recall that the imitation responses were also fitted with a logistic function to derive imitation gradiency; the resulting predicted curve shown in Figure *4.5* further confirms that participants were generally gradient in their imitation patterns.

Figure 4.5
Produced VOT across Stimulus VOT in the Imitation Task

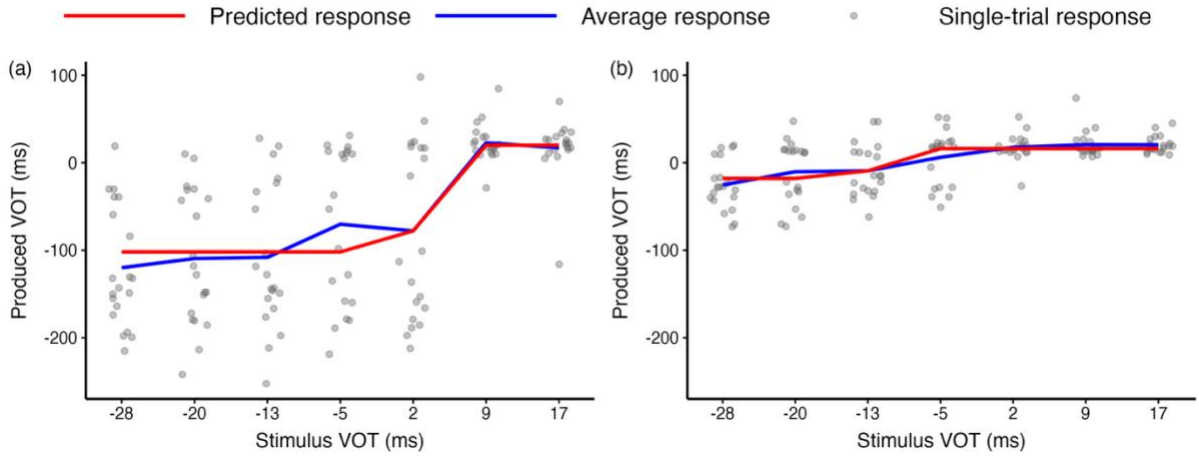


Note. The black line represents the grand average of participants' produced VOTs across stimuli with different VOT values, while the red line represents the fitted curve derived from the logistic function. Grey lines show individual participants' produced-VOT curves.

Participants also differed in their consistency of VOT production when the same stimulus was presented during the imitation task, as illustrated by two representative participants in Figure *4.6*.

Figure 4.6

Two Representative Participants in the Imitation Task: (a) Low Consistency and (b) High Consistency



Note. The blue line represents the average of each participant's observed imitated VOTs across stimulus VOT values, while the red line represents the predicted values from the fitted logistic function. Each dot indicates an individual trial imitation response. The same y-axis range is used for both panels to facilitate comparison.

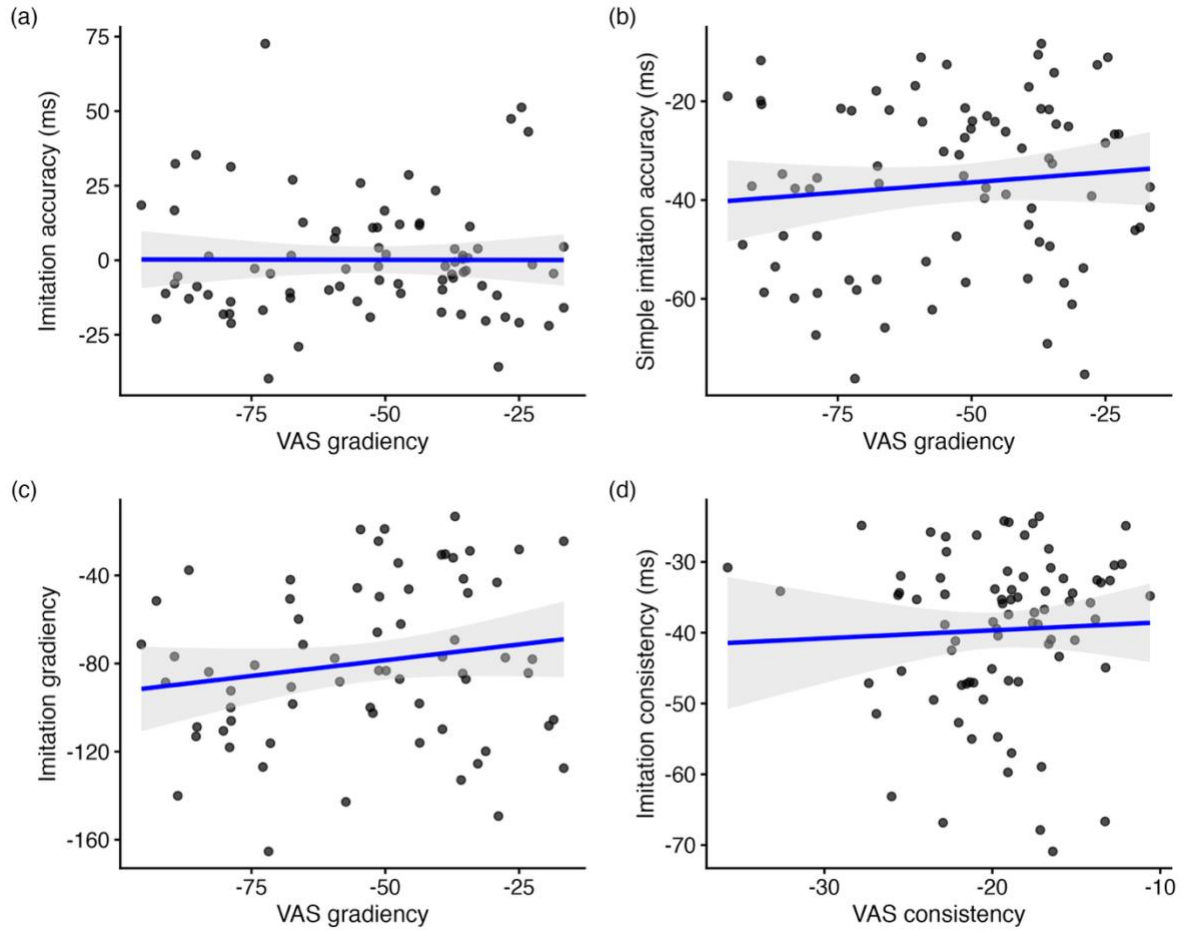
Next, we report Spearman's correlations among the measured variables. Table A4.1 presents the Spearman correlation coefficients (ρ) and corresponding Bayes factors (*BF*). We focus here on the meaningful relationships among key variables. Recall that we extracted four imitation measures: imitation accuracy, simple imitation accuracy, imitation gradiency, and imitation consistency. Results showed a positive correlation between simple imitation accuracy and imitation gradiency, $\rho = .59$, $BF = \infty$, with decisive evidence (Figure A4.1a). This suggests that participants who imitated the stimuli more closely (without accounting for baseline differences) tended to exhibit shallower slopes in their produced VOTs across stimulus VOTs, indicating greater imitation gradiency. A positive correlation was also found between simple imitation accuracy and imitation consistency, $\rho = .44$, $BF = 493.88$, showing decisive evidence (Figure A4.1b). This means that participants who imitated more closely to the prompt stimuli also produced more consistent VOTs across repetitions of the same stimulus. As expected, simple imitation accuracy was strongly correlated with imitation accuracy, $\rho = .59$, $BF = \infty$, confirming that both measures captured overlapping aspects of

imitation precision (Figure A4.1c). Furthermore, a positive correlation emerged between auditory acuity and articulatory ability, $\rho = .33$, $BF = 12.45$, with strong evidence (Figure A4.1d). Listeners with greater temporal sensitivity in auditory processing also tended to produce speech at a faster rate, reflecting better articulatory control.

Now we turn to our main research question: Do listeners who are more sensitive to within-category differences in speech perception also produce speech sounds with higher accuracy? As shown in Figure *4.7*a–c, the answer is clearly No. The correlations between VAS gradiency and imitation accuracy, $\rho = .02$, $BF = 0.17$, as well as simple imitation accuracy, $\rho = .07$, $BF = 0.21$, were weak and showed moderate evidence in favor of the null hypothesis. When considering imitation gradiency, the relationship was also weak, $\rho = .16$, $BF = 0.41$, and the Bayesian evidence was inconclusive, providing neither support for the null nor for the alternative hypothesis. For the second research question: whether listeners who are more consistent in speech perception also produce speech sounds more consistently, the answer is also No, as shown in Figure *4.7*d, $\rho = .12$, $BF = 0.30$, providing moderate evidence for the null.

Figure 4.7

Scatterplots Showing the Relationships between VAS Gradiency and (a) Imitation Accuracy, (b) Simple Imitation Accuracy, and (c) Imitation Gradiency, as well as (d) the Correlation between VAS Consistency and Imitation Consistency

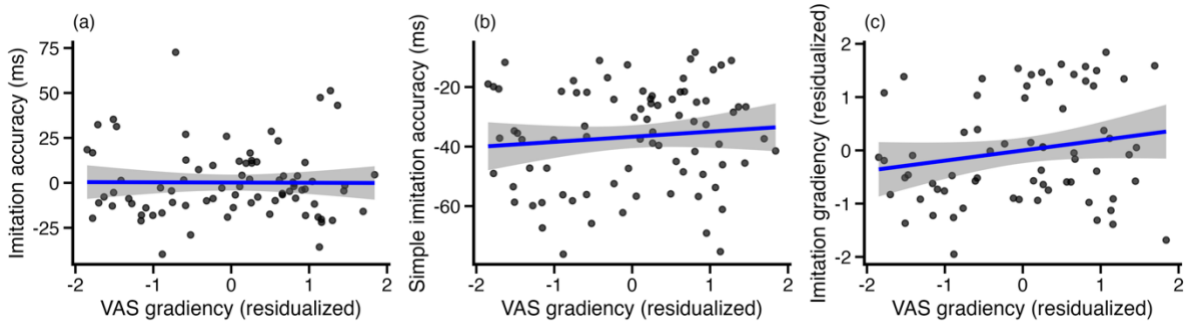


Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

As mentioned in Chapter 1, shallow VAS slopes combined with low consistency do not reflect true gradiency but rather noisy encoding. To account for this, we removed the variance in VAS gradiency explained by consistency and used the resulting residuals in all subsequent analyses (see Section 3.2.10). The same residualization procedure was applied to imitation gradiency. We then re-examined the relationship between the residualized VAS gradiency and imitation performance. Figure 4.8 presents the results. Overall, the correlations remained weak. There were very weak relationships between residualized VAS gradiency

and imitation accuracy, $\rho = .01$, $BF = 0.17$, as well as between residualized VAS gradiency and simple imitation accuracy, $\rho = .07$, $BF = 0.20$, with moderate Bayesian evidence supporting the null hypothesis. The relationship between residualized VAS gradiency and residualized imitation gradiency was also weak, $\rho = .17$, $BF = 0.46$, and the Bayesian evidence was inconclusive.

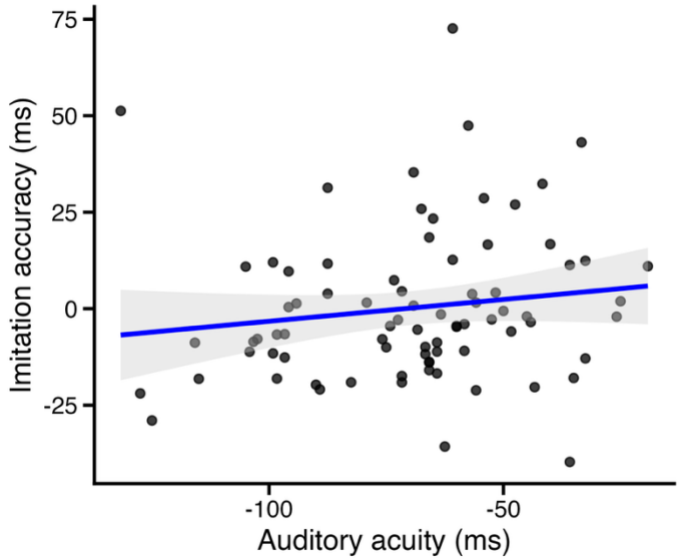
Figure 4.8
Scatterplots Showing the Relationships between Residualized VAS Gradiency and (a) Imitation Accuracy, (b) Simple Imitation Accuracy, and (c) Residualized Imitation Gradiency



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

Finally, the hierarchical regression analyses showed that VAS gradiency did not explain significant additional variance in Step 2 for imitation accuracy, simple imitation accuracy, or imitation gradiency (all $\Delta R^2 = 0.00–0.01$, all $p > .407$). Similarly, VAS consistency did not account for significant additional variance in Step 2 for imitation consistency, $\Delta R^2 = 0.01$, $p = .940$. Full statistical results are reported in Tables A4.2–A4.5 . Among the four models, the only significant predictor in Step 1 was auditory acuity, $\beta = 0.35$, $SE = 0.17$, $t = 2.05$, $p = .045$, when predicting imitation accuracy, indicating that participants with higher auditory acuity produced speech sounds more accurately (Figure *4.9*).

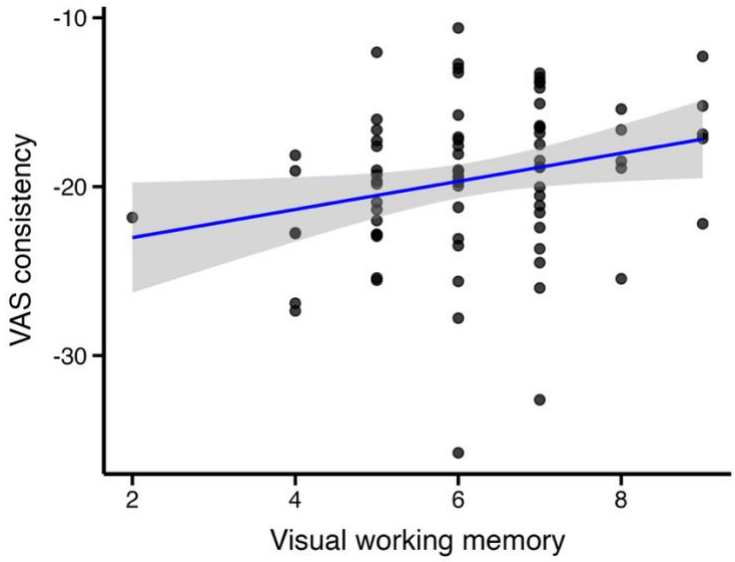
Figure 4.9
Scatterplot Showing the Relationship between Auditory Acuity and Imitation Accuracy



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

Recall that we also included a visual working memory task, which is nonlinguistic and nonauditory, to test whether VAS consistency reflects a more general individual trait rather than a speech-specific ability. There was a positive correlation between VAS consistency and visual working memory, $\rho = .30$, $BF = 5.79$ (Figure *4.10*), showing moderate evidence for a relationship between the two. This pattern suggests that consistency may, at least in part, reflect a broader cognitive trait related to task engagement and attention, rather than a speech-specific ability per se.

Figure 4.10
Scatterplot Showing the Relationship between Visual Working Memory and VAS Consistency



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

4.4 Discussion

In this study, we examined the relationship between speech perception and speech production at the subphonemic level for the Spanish /b/–/p/ contrast, focusing on fine-grained acoustic differences. Our goal was to test whether theoretical frameworks that posit a perception–production link can be extended to the subphonemic level. The answer appears to be No. Listeners who showed greater within-category sensitivity and consistency in speech perception did not demonstrate higher imitation accuracy or consistency in production, respectively. Most relationships showed moderate Bayesian evidence in favor of the null hypothesis, whereas the relationship between VAS gradiency and imitation gradiency was inconclusive. However, when considering other relevant variables in the regression analysis, we observed a positive relationship between auditory acuity and imitation accuracy. Importantly, this study addressed several limitations of prior work by employing a rigorous, tightly matched design, including presenting identical stimuli across perception and imitation

tasks, using a VAS measure to capture subtle perceptual gradiency, and implementing an explicit imitation task while controlling for other individual-difference variables. Even under these optimal conditions, the link between perception and production remained weak, suggesting that this relationship, if it exists, is quite limited at the subphonemic level.

For our first research question, whether perceptual gradiency predicts imitation accuracy, our findings suggest that frameworks such as the Interactive Alignment and Simulation Theories of speech imitation (Gambi & Pickering, 2013; Garrod & Pickering, 2004; Pickering & Garrod, 2004), which posit automatic priming of shared representations across linguistic levels, may not extend directly to the subphonemic domain. These models successfully explain imitation at higher linguistic levels (e.g., phonological or lexical), but the fine-grained acoustic alignment required for subphonemic imitation (such as small VOT differences) may depend on distinct, less automatic, or more motor-specific mechanisms.

It is important to note that our study tested only subphonemic variation in the Spanish /b/–/p/ contrast defined by VOT, and gradiency is known to be contrast-specific (e.g., Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; H. Kim et al., 2025d; Myers et al., 2024b; cf. Bidelman et al., 2026). Thus, our findings related to gradiency may not generalize to other contrasts. In fact, a recent study (Kuilin et al., 2025) found that listeners who exhibited greater gradiency when perceiving the English sibilant fricative contrast /s/–/ʃ/, rather than greater consistency or articulatory ability, achieved higher accuracy when imitating the novel Mandarin voiceless alveolo-palatal fricative /ɕ/. This finding suggests that perceptual gradiency may facilitate more precise articulatory planning and production. Listeners with higher perceptual gradiency may represent within-category acoustic variation more precisely, allowing them to detect finer discrepancies between the L2 target and their L1 production norms, thereby enabling more accurate imitation. However, we did not

observe a corresponding relationship for the temporal cue of VOT in L1 Spanish, suggesting that perception–production links may emerge for different cues, such as spectral shape, rather than for very fine-grained temporal VOT differences.

The general imitation pattern in the present study, particularly the relatively flat portion of the curve between −28 and −13 ms in Figure *4.5*, aligns with the findings of Mitterer and Ernestus (2008). Although their work used a shadowing task rather than explicit imitation, they demonstrated that phonetic imitation is primarily driven by phonological distinctions. Mitterer and Ernestus (2008) manipulated Dutch /b/ and /d/ tokens to create three versions: no prevoicing (0 ms), canonical prevoicing (−38 ms), and lengthened prevoicing (−64 ms). Participants reliably reproduced the contrast between prevoiced and non–prevoiced stops, but they did not imitate the finer difference between canonical and lengthened prevoicing. Because only the former contrast is phonologically meaningful in Dutch, the results suggest that shadowing is more sensitive to phonological than to purely phonetic variation. It is worth noting that prevoicing in Dutch is typically shorter than in Spanish, where voiced stops often exhibit longer negative VOT values (Lisker & Abramson, 1964). Nevertheless, our findings similarly suggest that, for Spanish, the relationship between speech perception and production operates largely at an abstract phonological level.

While the present study focused on the primary cue of voicing for bilabial consonants, VOT, future work should also examine secondary cues. Recent findings show that when Mandarin speakers imitate lexical tones, imitation of F_0 range tends to be relatively categorical, whereas imitation of duration is more gradient (Zhang et al., 2023). This pattern suggests that imitation may depend on cue weighting, or possibly on inherent cue-specific constraints, such as the greater difficulty of continuously imitating F_0 compared to duration. Although these findings come from suprasegmental contrasts, an important next step is to test

whether similar cue-dependent differences in gradiency emerge at the segmental level, and whether they relate to individual differences in speech perception.

Another promising direction is to test whether listeners with no prevoicing experience imitate prevoicing more gradiently. Evidence from lexical tone imitation suggests that English speakers with no lexical-tone experience show more gradient F_0 imitation than Mandarin speakers (Zhang et al., 2025), raising the possibility that phonetic experience modulates imitation gradiency across cue types.

Additionally, future work could test whether gradiency and consistency influence listeners' ability to modify their productions in training paradigms such as visual-feedback articulatory training, which provides learners with visual displays of their productions relative to a target (e.g., vowel-space plots). Although such training has been shown to support the acquisition of L2 sounds (Kartushina & Martin, 2019, 2025), the factors that drive individual differences in training success remain only partially understood. Identifying whether perceptual gradiency and consistency contribute to these differences would clarify why some learners benefit more than others and could help inform the design of more effective L2 instructional methods.

We also found a weak relationship between perceptual consistency and production consistency, which was somewhat unexpected given the strong correlations observed between consistency measures across contrasts (e.g., H. Kim et al., 2025e) and across modalities (e.g., H. Kim et al., 2025d). This dissociation suggests that perceptual and production consistency may rely on at least partly distinct neural and cognitive mechanisms.

Models of speech motor control (Hickok et al., 2011; Houde & Nagarajan, 2011; Tourville & Guenther, 2011) propose that perception guides production through an auditory

feedback loop, and that sharper auditory representations can lead to more precise speech output (Perkell, 2012). While such models are supported by vowel data showing that individuals with better auditory discrimination produce less variable vowels (Franken et al., 2017; Perkell et al., 2004, 2008), our findings suggest that this link does not generalize to consonantal VOT contrasts. Vowels provide salient and continuous acoustic cues, whereas subtle temporal differences in VOT are inherently harder to imitate consistently, especially with step sizes as small as 7–8 ms. Thus, the weak perception–production relationship observed here likely reflects the greater motor and temporal demands of this type of subphonemic imitation.

We also explored additional potential predictors of imitation ability, including articulatory ability, task difficulty, openness to experience, attention switching, and musical perception skill. None showed strong effects. Prior studies linking these variables to imitation typically examined word- or accent-level imitation (Myers et al., 2024a; Yu et al., 2013), whereas the present study targeted subphonemic imitation. These results therefore suggest that such individual-difference factors play a smaller role when the task involves fine-grained, low-level acoustic imitation.

We found that auditory acuity was the only variable that reliably predicted imitation accuracy in the regression analysis. Recall that auditory acuity is a nonlinguistic, domain-general measure derived from a temporal order judgment task with pure tones, tapping sensitivity to brief timing differences. This temporal resolution is conceptually similar to the information required to perceive VOT, which is also defined by fine-grained temporal cues. Previous work reported that auditory acuity predicted perceptual gradiency in Spanish speakers (Kapnoula & Samuel, 2024). Although we did not observe this relationship in the present study, individuals with higher auditory acuity were better at imitating the speech

sounds, once baseline production tendencies were taken into account. This suggests that auditory acuity may play a more prominent role in the sensorimotor mapping required for imitation than in explicit perceptual judgments. More broadly, this "good ear" advantage has been repeatedly linked to successful second-language acquisition (Kachlicka et al., 2019; Kempe et al., 2014; Saito et al., 2021, 2022; Shao et al., 2023), and the present findings extend this work by showing that individuals with sharper auditory encoding are also more accurate speech sound imitators.

Taken together, the results indicate that individual differences in phonetic imitation gradiency are not driven by speech perception gradiency, nor by broader cognitive, personality, or musical factors. This raises an open question: why are some individuals better at subphonemic imitation than others? Our findings suggest that auditory acuity may play a role, but further work is needed to identify the mechanisms that give rise to these individual differences.

In summary, the present study provides clear evidence that the relationship between speech perception and production does not directly extend to the subphonemic domain. Even when perception and imitation were tightly matched in stimuli, listeners who were more gradient or consistent in perception did not show correspondingly higher precision in production. These findings suggest that perception–production links observed at phonological or lexical levels may weaken or disappear for fine acoustic details such as VOT. Instead, subphonemic imitation likely depends on more motor-specific or temporally constrained mechanisms. Together, these results contribute to refining theories of speech perception and production by identifying the limits of their interaction.

5 General Discussion

Throughout this dissertation, we have seen that listeners vary in the degree of flexibility and stability they exhibit in subphonemic speech perception, reflected in the constructs of gradiency and consistency, respectively. These two constructs appear to be largely orthogonal. While recent studies focusing on English have begun to illuminate the sources of gradiency and some functional roles of gradiency and consistency, the present dissertation extends this line of work by examining the malleability of gradiency. In addition, it investigates the functional roles of these constructs in initial lexical activation and speech perception flexibility, and extends this inquiry to speech imitation in the production domain. As discussed below, the findings from this dissertation also shed light on the sources of consistency and gradiency and their characteristics. This dissertation centers on Spanish, a language whose VOT range differs from that of the most extensively studied language, English. Some chapters also incorporate additional languages, including Basque and English, enabling a broader examination of L2 experience and cross-linguistic influences on gradiency. The following sections summarize the main findings of the dissertation before turning to their broader implications.

5.1 Summary of Results

Experiment 1 (Chapter 2) demonstrated that gradiency is malleable when listeners are exposed to different distributional statistics. Specifically, listeners became more gradient when the distributional variance was wider. This malleability was more evident when two languages shared similar VOT ranges (i.e., Spanish and Basque) than when the two languages differed substantially in their VOT ranges (i.e., Spanish and English).

Experiment 2 (Chapter 3) revealed that consistency facilitated initial lexical activation in both L1 Spanish and L2 English, whereas gradiency facilitated speech perception

flexibility in L1 Spanish. Gradiency was positively correlated across L1 and L2 among listeners with higher L2 proficiency, while consistency showed a positive overall correlation across languages.

Experiment 3 (Chapter 4) indicated that the relationship between speech perception and production was weak for both gradiency and consistency. Listeners with higher perceptual gradiency or consistency did not necessarily imitate speech sounds more accurately or more consistently.

5.2 Malleability of Gradiency

Examining the malleability of gradiency is theoretically important because it speaks to whether gradiency reflects a stable individual trait or a dynamic property shaped by recent input statistics. If gradiency can be modified through passive exposure, this would suggest that phonetic category structure remains flexible even in adulthood. Conversely, if gradiency is resistant to short-term manipulation, this would support the view that it primarily reflects accumulated long-term experience. Clarifying the extent and limits of this malleability therefore has implications for models of speech perception and for understanding individual differences in phonetic processing.

In Experiment 1, we tested whether distributional statistics influence gradiency using both categorization slope and competitor fixation measures. Based on the competitor fixation measure, which we argue provides a more sensitive index of gradiency, we found that gradiency can be modulated by passive exposure to distributional input, consistent with the conclusions of Kapnoula (2016). Importantly, this effect was more pronounced when the VOT overlap between languages was relatively large or when the overall combined VOT range was constrained (e.g., Spanish–Basque). By contrast, when the combined VOT range is very wide (e.g., Spanish–English), listeners may be less sensitive to the experimentally manipulated variance. One possibility is that when the overall acoustic range becomes too

broad, listeners no longer treat the manipulated distribution as informative. Alternatively, extensive long-term experience with this broader acoustic space may dilute the impact of local distributional statistics. This pattern may also reflect the influence of L1-based priors, an issue that will be discussed further in Section 5.6.

Regardless of the mechanism, our findings underscore the importance of considering language-specific phonetic characteristics when measuring gradiency and have theoretical implications for models of speech perception. For example, while the TRACE model assumes gradient activation of lexical representations (McClelland & Elman, 1986), it does not explicitly formalize how cross-linguistic differences in VOT range or phonetic distribution may modulate the degree of gradiency observed. More broadly, models of spoken word recognition typically allow graded activation during lexical competition but treat the mapping between acoustic input and phoneme representations as given, rather than modeling how language-specific distributional properties shape phonetic category structure.

Experiment 1 also offers methodological insights. Following Clayards et al. (2008), we used both categorization slopes based on mouse-click responses and competitor fixation measures in a VWP to assess distributional learning effects. However, we observed divergent results from these two measures within each group of participants. As discussed in Chapter 2, both measures have limitations in capturing gradiency compared to the VAS slope. In particular, categorization slope may reflect consistency rather than gradiency (Honda et al., 2024a). This interpretation is consistent with the present findings: the Spanish–English group, which had lower overall proficiency than the Spanish–Basque group, exhibited shallower slopes overall. We interpret this pattern as reflecting lower consistency rather than greater gradiency, given that lower proficiency is more plausibly associated with the former.

The malleability of gradiency is also related to characteristics of the VAS design, which may inform future research. In the present study, as in most previous work, the VAS

allowed participants to respond at their own pace and adjust their responses freely. However, some studies have included RT as an additional measure and reported slower responses to ambiguous steps (e.g., Bidelman et al., 2025; Rizzi & Bidelman, 2024), suggesting increased processing demands near category boundaries. Future research could examine whether instructing participants to respond under speeded conditions alters observed gradiency, as time pressure may encourage greater reliance on early acoustic encoding. Additionally, it remains unclear whether stimulus range, while maintaining small step sizes, influences observed gradiency. Previous work has shown that category boundaries can shift depending on the range of stimuli presented (Rosen, 1979). Expanding the stimulus range while maintaining the small step sizes may also influence gradiency, paralleling findings that increased acoustic variability promotes more gradient categorization (e.g., Clayards et al., 2008; Gür & McMurray, 2025).

5.3 Functional Roles of Gradiency and Consistency

Experiment 2 shows that gradiency facilitates speech perception flexibility, whereas consistency supports initial lexical activation. The weak relationship between consistency and speech perception flexibility, together with evidence that gradiency predicts flexibility, suggests that consistency and gradiency serve distinct functions in speech processing. In the VWP paradigm used in Experiment 2, flexibility (indexed by recovery rate and recovery latency) reflects listeners' ability to revise an initially misleading interpretation once disambiguating information becomes available. One possibility is that recovery from misleading input relies less on stable category commitments and more on listeners' sensitivity to fine-grained acoustic variation and their ability to maintain multiple lexical candidates during processing (Kapnoula et al., 2021). These processes are more naturally captured by gradiency than by consistency. In contrast, consistency reflects the stability of phonetic categorization across repeated stimuli and may primarily support early, input-

compatible lexical activation rather than late-stage revision. Consequently, lower consistency does not necessarily hinder recovery from misleading input: listeners with less stable categorization may initially activate lexical representations less strongly, but this weaker activation neither precludes nor facilitates flexible adjustment when disambiguating information becomes available. Taken together, these findings suggest that initial lexical activation and speech perception flexibility rely on dissociable mechanisms. These findings are consistent with interactive models of spoken word recognition in which phonetic encoding influences both the strength of initial lexical activation and the subsequent resolution or revision of lexical competition. However, the present results suggest that these processes may rely on partially dissociable representational properties.

In addition, we observed that listeners with varying levels of L2 proficiency relied more on top-down lexical information in L2 English than in L1 Spanish, possibly due to less robust acoustic-phonetic representations of the relevant /b/–/p/ contrast. Notably, using the same stimuli and experimental paradigm, Kapnoula et al. (2021) reported that monolingual English listeners exhibited a similar pattern when processing English stimuli, comparable to the pattern observed in our L1 Spanish participants. Moreover, language proficiency shapes the development of gradiency in the L2, influencing its alignment with the L1. Together, these findings suggest an additional constraint on models of speech perception beyond language-specific characteristics (e.g., VOT range and overlap). Specifically, they imply that the mapping from phonetic detail to lexical activation is not fixed but experience-dependent. Many influential models of spoken word recognition were developed primarily with monolingual listeners in mind and do not explicitly specify how L2 experience shapes the use of subphonemic information. Incorporating experience-dependent mechanisms into such models could help explain when and why the functional roles of gradiency and consistency observed in the L1 extend, or fail to extend, to the L2.

Beyond its implications for speech perception models, the finding that speech perception consistency corresponds across L1 and L2 even in lower-proficiency listeners, whereas cross-language similarity in gradiency emerges only among higher-proficiency listeners, may also reflect a developmental sequence in the acquisition of L2 phonetic representations. One possible (albeit speculative) interpretation is that L2 consistency may develop earlier than L2 gradiency within the same individuals. Consistency likely depends on accumulating sufficient input to establish robust phonetic categories (potential sources of consistency will be discussed in Section 5.4). Once these mappings are stable, finer gradations within categories can begin to influence perception, giving rise to gradiency. In this view, consistency provides the foundational stability upon which gradiency operates, supporting flexible and efficient speech perception. Consistent with this proposal, recent longitudinal work tracking school-aged children appears to indicate that consistency improves over the first several years of development, whereas gradiency appears to unfold over a more extended time span (Kutlu et al., 2026). Future longitudinal research in adults will be essential to test this proposed developmental sequence directly.

In Experiment 3, while we demonstrated that speech imitation can be gradient, we observed only weak associations between perceptual and production measures of both gradiency and consistency. In other words, listeners with higher gradiency or consistency did not necessarily exhibit stronger speech alignment or convergence. These findings challenge strong versions of speech perception–production models that assume a tightly coupled relationship between the two systems. Importantly, our study targeted very fine-grained subphonemic variation (small VOT step sizes) and therefore does not contradict previous work demonstrating perception–production links at higher representational levels, such as phonological or lexical processing (e.g., Goldinger, 1998; Pardo et al., 2018). While our findings may not generalize to more perceptually salient acoustic dimensions, such as vowel

quality differences, the present results place important constraints on models that posit a uniform perception–production linkage, suggesting instead that the strength of this relationship may depend on representational level and acoustic granularity.

5.4 Sources of Consistency and Gradiency

As discussed throughout this dissertation, speech perception consistency confers advantages across multiple domains of speech perception and language learning. However, its underlying sources remain unclear. Broadly, three possibilities can be considered, corresponding to three levels of processing: (1) early sensory encoding, (2) the mapping between acoustic cues and phoneme categories, and (3) downstream decision-level processes.

Experiment 2 provides several constraints on these possibilities. Consistency facilitated initial lexical activation but did not predict speech perception flexibility. This pattern makes a purely decision-level account less likely, because initial lexical activation reflects early processing before the point of disambiguation, occurring at around 450 ms. At the same time, consistency is unlikely to arise solely from early sensory encoding. As discussed earlier, previous work has shown that consistency measured using the VAS task is not associated with the consistency of the frequency-following response (FFR; Honda et al., 2024b). This dissociation suggests that consistency does not primarily reflect variability in early acoustic encoding.

Taken together, the present findings are most consistent with the interpretation that consistency reflects the stability of the mapping between continuous acoustic input and phoneme categories. Under this account, listeners with higher consistency map acoustic cues onto phonological categories more reliably, which strengthens the influence of early acoustic evidence on lexical activation. In contrast, speech perception flexibility involves revising an initial interpretation once disambiguating information becomes available. This later process depends more on how listeners update lexical activation in response to new input than on

how stable the initial cue-to-category mapping is. As a result, consistency may influence the strength of early lexical activation but not the ability to revise interpretations during speech perception.

Another factor to consider is that consistency may be modulated by top-down influences. For example, a previous study found that listeners were more consistent when processing words than nonwords, suggesting that lexical knowledge may contribute to stabilizing phonetic categorization through a feedback mechanism (Carraturo et al., 2025). However, the present findings provide limited support for this interpretation. First, consistency predicted early lexical activation but not speech perception flexibility, even though flexibility plausibly places greater demands on lexical knowledge and top-down support than early activation. If consistency were primarily modulated by top-down factors, one might expect its influence to extend to recovery from garden-path misperceptions, particularly in L1 Spanish, where lexical representations are well established. However, no such effect was observed in the present study.

Second, if consistency were primarily driven by top-down feedback, listeners with higher L2 proficiency, who presumably possess stronger linguistic knowledge, might be expected to show higher consistency. However, the present results provide no clear evidence for a relationship between L2 proficiency and L2 consistency, and therefore do not support strong top-down effects in L2 perception. That said, it is worth noting that most participants were Spanish residents with limited immersive experience in English and primarily learned English in L2-speaking environments, where they frequently encountered fellow learners and teachers producing difficult sounds in noncanonical ways (Eger & Reinisch, 2019). Future research examining listeners with more extensive L2 experience will be needed to clarify the role of proficiency and language-specific contrast properties in shaping the functional roles of consistency. In addition, the Bayesian evidence for the correlation was inconclusive, and the

VAS task used nonwords, which may have reduced the influence of lexical knowledge. Hence, while our findings place meaningful constraints on this top-down account, further work will therefore be needed to clarify the extent to which higher-level linguistic knowledge contributes to individual differences in speech perception consistency.

Based on the arguments above, we suggest that speech perception consistency likely stems from the stability of cue-to-category mappings, with limited influence from top-down processes. However, another possibility must also be considered: consistency may reflect differences in general effort or attention. Because consistency appears to confer advantages across multiple domains, one might speculate that listeners with higher consistency are simply more attentive or exert greater effort during the task. Supporting this possibility, previous work has shown that listeners with higher speech perception consistency also exhibit higher consistency in other tasks, including visual VAS and 2AFC tasks (Honda et al., 2024a; H. Kim et al., 2025d). Furthermore, several studies have reported broad advantages of consistency in speech perception and reading (Carraturo et al., 2025; H. Kim et al., 2025a–d, 2025f; Fuhrmeister et al., 2023; Honda et al., 2024a, 2024b; Myers et al., 2024b; Rizzi & Bidelman, 2025; Kutlu, 2025).

However, we argue that a purely effort- or attention-based explanation is unlikely for several reasons. First, Experiment 2 showed that listeners with higher speech perception consistency exhibited stronger initial lexical activation, but not enhanced speech perception flexibility. This selective effect suggests that consistency does not confer uniform benefits across all aspects of processing, thereby weakening a purely effort-based explanation. If consistency merely reflected greater effort or attention, one would expect improvements across multiple processing stages, including flexibility and other measures examined throughout the dissertation. Table *5.1* summarizes the Spearman correlations between potential source variables and consistency (as well as gradiency; discussed below) in L1

Spanish across experiments. Overall, the relationships between these variables and consistency were generally weak, with Bayesian evidence ranging from inconclusive to moderate support for the null. Taken together, this pattern suggests that consistency cannot be explained solely by effort.

Table 5.1
Summary of Spearman's Correlations between Potential Source Variables and Gradiency and Consistency in L1 Spanish, with BF in Parentheses

	N	Gradiency	Consistency
Auditory acuity (Experiment 3)	80	.03 (0.18)	.09 (0.23)
Inhibitory control (Experiment 2)	70	−.25 (2.02)	−.08 (0.18)
Visual working memory (Experiment 2)	70	.01 (0.27)	.14 (0.30)
Visual working memory (Experiment 3)	80	−.17 (0.50)	**.30 (5.79)**
Musical training duration (Experiment 2)	70	.15 (0.56)	.14 (0.28)
Musical perception skills (Experiment 3)	80	−.14 (0.38)	−.03 (0.18)

Note. Correlations with at least moderate evidence, based on Bayes factors, are shown in bold.

From Table *5.1*, the only variable that showed a relationship with consistency was visual working memory in Experiment 3. Specifically, listeners with higher speech perception consistency also demonstrated higher visual working memory, a domain not directly related to speech perception. While this relationship may reflect differences in general effort or attention, it may also be explained by the possibility that listeners with stronger working memory are better able to maintain stimulus representations or task-related information during the experiment, which could indirectly contribute to more stable responses. This speculation does not contradict the cue-to-category mapping account discussed above. However, this interpretation remains speculative, and the same relationship was not observed in Experiment 2. To more directly examine the relationship between

working memory and consistency within the auditory domain, future studies should test whether auditory working memory is associated with speech perception consistency.

Importantly, we do not argue that effort and attention play no role in consistency. Rather, a certain degree of effort and attention is likely required for listeners to perform the VAS task consistently (as in any task). However, these factors are unlikely to be the primary sources of individual differences in speech perception consistency. Even when listeners exert similar levels of effort or attention, differences in the stability of cue-to-category mappings may still lead to differences in speech perception consistency.

Compared to consistency, the source of gradiency is clearer. Recall that previous work has suggested that gradiency reflects early perceptual encoding, based on evidence linking VAS gradiency to the N1 ERP component (Kapnoula & McMurray, 2021). Although the studies in this dissertation did not directly examine the neural sources of gradiency, our findings provide indirect support for this interpretation. In Experiment 1, gradiency appeared to be malleable through exposure to distributional statistics. At the same time, previous work suggests that gradiency is not strongly modulated by explicit feedback or training (Giovannone et al., 2024; Kapnoula, 2016; Nguyen et al., 2024). This pattern is informative because early perceptual encoding is typically tuned through exposure to the statistical structure of the input, whereas explicit feedback more often influences later decision or categorization processes. Taken together, these findings are therefore consistent with the view that gradiency reflects early perceptual encoding processes.

Converging with earlier findings (e.g., Kapnoula et al., 2017; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; Kong & Edwards, 2016), we found that general cognitive factors such as inhibitory control and working memory were not reliably associated with gradiency (Table *5.1*). In addition, Experiment 2 showed that L1 and L2 gradiency were aligned only among listeners with higher L2 proficiency. Taken together, these findings

suggest that gradiency may reflect language-specific experience rather than domain-general perceptual abilities. This interpretation aligns with recent findings showing that linguistic diversity and cumulative language exposure enhance gradiency (Kapnoula & Samuel, 2024; Kutlu et al., 2024; McMurray et al., 2018)

Overall, the present findings constrain theoretical accounts of speech perception consistency by locating it at the cue-to-category mapping level, while working memory may play a limited supporting role. Gradiency, on the other hand, is likely related to early perceptual encoding and appears to be fine-tuned by language experience rather than driven by general cognitive factors.

5.5 Characteristics of Gradiency and Consistency

As discussed in Chapter 1, it is important to take consistency into account when measuring gradiency at the behavioral level, because inconsistent trial-to-trial responses can produce a shallow VAS slope even when a listener is fundamentally categorical. Recent studies have begun to incorporate consistency when estimating gradiency (e.g., Kutlu et al., 2026; Sorensen et al., 2024), thereby strengthening the construct validity of gradiency measures. Similarly, throughout this dissertation, we examined gradiency while controlling for consistency through residualization. Across the /b/–/p/ contrast examined in this dissertation, the relationship between unresidualized gradiency and consistency measured by the VAS was generally weak (see Table *5.2* for a summary). The weak association helps explain why residualizing gradiency with respect to consistency did not alter the results across experiments. Nonetheless, we maintain that accounting for consistency remains methodologically important when interpreting gradiency measures, particularly in light of previous studies reporting stronger associations between the two constructs (e.g., Fuhrmeister et al., 2023; Honda et al., 2024a).

Table 5.2
Summary of Spearman's Correlations between Unresidualized Gradiency and Consistency, with BF in Parentheses

	N	Correlation between gradiency and consistency
Experiment 2 (Spanish)	70	.18 (0.80)
Experiment 2 (English)	56	.08 (0.35)
Experiment 3	80	−.10 (0.24)

As discussed throughout the dissertation, previous findings suggest that gradiency and consistency differ in their degree of generality and stability (e.g., Kapnoula et al., 2021; H. Kim et al., 2025d). Gradiency appears to be more contrast- and modality-specific, varying across linguistic contexts and tasks. In contrast, consistency behaves more like a stable trait. Several findings in the present dissertation support this distinction. In Experiment 2, L1 Spanish and L2 English consistency were positively correlated, whereas gradiency showed no overall cross-language association, except among higher-proficiency L2 listeners. In addition to contrast and modality, this finding suggests that gradiency is also more language-specific than consistency. However, not all findings fully conformed to this proposed distinction. For example, neither gradiency nor consistency showed robust perception–production associations in Experiment 3.

Taken together, the current results and prior findings support the view that consistency reflects a relatively stable, trait-like dimension, whereas gradiency is more context-dependent and sensitive to task or contrast characteristics. However, the present dissertation suggests that this distinction should not be treated as absolute. Both constructs likely interact with individual differences and language experience. Importantly, existing models of spoken word recognition generally do not formalize stable trait-like variation separately from contrast-specific representational precision. Explicitly incorporating this distinction may therefore offer a valuable direction for future theoretical development.

5.6 Cross-Linguistic Generality of Gradiency and Consistency

Most previous studies on gradiency have been conducted in English, with many focusing specifically on VOT contrasts (e.g., Kapnoula et al., 2017, 2021; McMurray et al., 2002; Toscano et al., 2010). The present dissertation extends this literature to Spanish, which differs from English in two important ways. First, Spanish includes prevoicing. Second, Spanish and English differ in the distribution of the VOT cue. As illustrated in Figure *1.3*, Spanish /b/ shows greater variability than Spanish /p/ (Souganidis et al., 2024), whereas in English, /p/ shows greater variability than /b/ (Allen & Miller, 1999). In addition, the variability of Spanish /b/ is greater than that of English /p/, contributing to a broader overall VOT range in Spanish than in English. These differences raise two considerations for the present dissertation.

First, it may be inherently more difficult for listeners to exhibit gradiency in the prevoicing range. Prevoicing consists of low-amplitude periodic voicing that precedes the burst, whereas positive VOT is characterized by higher-amplitude aspiration with a clearer acoustic onset. For example, in the VAS data from Experiment 3 (Figure *4.2*), listeners generally showed relatively flat average curves in the prevoicing range, particularly between −28 and −13 ms, suggesting limited sensitivity to variation. Note that the VAS task requires small step sizes to adequately capture gradiency. For this reason, in all experiments the most extreme /b/ step (−35 or −28 ms) was still far from the prototypical /b/. Typical bilabial VOT values in Peninsular Spanish are approximately /b/ ≈ −80 ms and /p/ ≈ 16 ms (Souganidis et al., 2024). Covering the full prototypical range for both categories while also maintaining small step sizes would have required a very large number of stimuli, substantially increasing task duration. Despite this challenge, we argue that using smaller step sizes across a wider prevoicing range would be informative for testing whether gradiency can be observed in that

range. In contrast, English may present fewer practical constraints in this respect because the overall VOT range between /b/ and /p/ is smaller (/b/ ≈ 16 ms; /p/ ≈ 70 ms; Allen & Miller, 1999), although prevoicing may occasionally occur in English as well (Lisker & Abramson, 1964). This may also have implications for the speech perception–production link. As discussed in Chapter 4, one possible reason for the weak perception–production association is that the Spanish speech stimuli were relatively far from the prototypical /b/ region, which may have increased task difficulty. Importantly, despite our Spanish stimuli being less extreme than prototypical Spanish /b/, the VAS responses generally spanned the full scale, indicating that listeners still categorized the stimuli as /b/ at the lower end. This suggests that using a less prototypical /b/ in the VAS did not undermine the interpretability of our results.

In addition, the cross-linguistic differences in VOT distributions are relevant to Experiment 1, in which we tested the malleability of gradiency through distributional variance manipulations. Although we used a relatively wide VOT range (−65 to 45 ms), bringing the /b/ tokens closer to their prototypical values, the stimuli still did not span the full prevoicing range observed in natural Spanish speech. The stimulus range was shifted toward the /p/ side in order to place the category boundary near the center of the continuum, similar to the English version. In contrast, the English range (−30 to 80 ms) more fully covered the relevant distributional space. The extent to which these experimental distributions matched or mismatched listeners' prior expectations remains unknown. While our primary focus was on how distributional variance shapes gradiency, we did not explicitly model participants' prior expectations about category distributions. According to the ideal adapter framework (Kleinschmidt & Jaeger, 2015), speech perception reflects the integration of long-term prior distributions and current input statistics. Because it is not feasible to directly measure listeners' prior distributions, and participants themselves are unlikely to have explicit access to them, computational modeling could provide a principled approach for estimating how

prior expectations interact with experimentally manipulated distributions. This approach could also test one of our interpretations of the distributional learning results. Specifically, the larger learning effect observed when the VOT overlap range was small may reflect hierarchical inference processes. When the VOT distributions of two languages differ substantially, distributional learning may become more difficult because listeners cannot easily rely on L1-based priors to interpret L2 input (Pajak et al., 2016). In such cases, greater divergence between prior expectations and incoming statistics may increase the demands on perceptual updating. Although hierarchical inference models do not explicitly incorporate variables such as proficiency, age of acquisition, or exposure, these factors likely influence the strength and stability of listeners' prior distributions and may therefore modulate the ease of updating.

Despite the acoustic differences between Spanish and English, we show that most findings on gradiency previously reported for bilabial contrasts in English can be extended to Spanish. Specifically, listeners exhibit generally gradient perception but with individual differences, gradiency reflects early sensory encoding, and its associations with nonlinguistic variables are weak. However, language-specific factors, such as the presence of prevoicing and differences in phoneme distributions, may modulate how gradient perception manifests. Notably, these differences may emerge even when examining a single bilabial voicing contrast. It is therefore plausible that greater cross-linguistic variation would be observed when considering additional phonetic contrasts or suprasegmental features. Cross-linguistic investigation is therefore essential, as we cannot assume that all languages converge on identical speech perception mechanisms.

5.7 Conclusions

The present dissertation demonstrates that subphonemic speech perception is multidimensional rather than governed by a single underlying construct. That is, at least two theoretically dissociable constructs, gradiency and consistency, shape how listeners encode, interpret, and act upon fine-grained acoustic information. These constructs differ in their sources, functional roles, and cross-linguistic generality, yet together they jointly contribute to the architecture of speech perception.

Beyond establishing a functional dissociation, the dissertation further shows that both language-specific characteristics and language experience constrain how these constructs operate. The malleability of gradiency depends in part on the distributional structure of the acoustic-phonetic space in the languages involved. In addition, cross-language alignment of gradiency emerges primarily among bilinguals with higher L2 proficiency, and these listeners may rely differently on top-down versus bottom-up cues across L1 and L2. Together, these observations indicate that the influence of subphonemic detail on lexical processing is shaped by both language structure and cumulative language experience. Accordingly, both the use of acoustic detail and its mapping to lexical activation are neither fixed nor uniform across languages, but vary as a function of language structure and experience.

Future models of spoken word recognition should distinguish between sensitivity to fine-grained acoustic detail and the stability of phonetic categorization, while also accounting for language-specific structure and experience. By integrating these dimensions, we move closer to a more comprehensive account of how listeners navigate subtle acoustic differences in speech. Although these differences may seem small, they can have substantial consequences, particularly in ambiguous or misleading everyday listening environments shaped by background noise, speaker accent, and coarticulation. In doing so, we recognize that speech perception is not merely precise or imprecise, but dynamically shaped by language structure and experience, giving rise to meaningful individual differences.

List of Figures

Figure 1.1 Visual Analog Scale (VAS) Task 6
Figure 1.2 Examples of Spanish VAS Data Showing Different Gradiency and Consistency of Four Participants 7
Figure 1.3 VOT Differences between English, Spanish, and Basque /b/–/p/ Contrasts 19
Figure 2.1 Number of Trials at Each VOT Value in the Narrow and Wide Distribution Conditions Used in Clayards et al. (2008) 23
Figure 2.2 *VWP Used in Experiments 1.1 and 1.2* 33
Figure 2.3 Mouse-click Responses of (a) Experiment 1.1 and (b) Experiment 1.2 40
Figure 2.4 Fixation Patterns of (a) Experiment 1.1 and (b) Experiment 1.2 41
Figure 2.5 Participant-by-Item Categorization Slope as a Function of Distributional Condition and Language in (a) Experiment 1.1 and (b) Experiment 1.2 42
Figure 2.6 Proportion of Fixations on Competitor Pictures by Distributional Variance and Language in (a) the SNBN and SWBW Groups, (b) the SNBW and SWBN Groups, (c) the SNEN and SWEW Groups, and (d) the SNEW and SWEN Groups 45
Figure 3.1 VWP Display 86
Figure 3.2 Order of Tasks for Each Group 91
Figure 3.3 Illustration of Minimum and Maximum Asymptotes, Slope, Crossover, and Pointwise Variability Extracted from the VAS Task 92
Figure 3.4 VAS Rating across VOT for (a) Spanish and (b) English 94
Figure 3.5 Examples of Spanish VAS Data Showing Different Gradiency and Consistency of Four Participants 95
Figure 3.6 Examples of English VAS Data Showing (a) Lower Gradiency and (b) Higher Gradiency in Two Participants 95
Figure 3.7 Average English VAS Data for (a) Low English Proficiency Group and (b) High English Proficiency Group 98
Figure 3.8 Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in (a) Spanish VWP and (b) English VWP 100
Figure 3.9 Predicted Accuracy, RT, and Proportion of Fixations on the Target as a Function of tDist for Voiced and Unvoiced Targets 103
Figure 3.10 Target Fixations as a Function of Time and tDist in (a) Spanish and (b) English VWPs 105
Figure 3.11 Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) /b/-Onset Words and (b) /p/-Onset Words 107
Figure 3.12 Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency 107
Figure 3.13 Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency for /b/-Onset Words 108
Figure 3.14 Proportion of Responses to Target, Competitor, Filler, and X as a Function of Distance from the Target (tDist) in the English VWP for (a) Listeners with Higher English Proficiency and (b) Listeners with Lower English Proficiency for /p/-Onset Words 108
Figure 3.15 Correlation between L1 and L2 Gradiency in (a) the Full Dataset, (b) High Proficiency Listeners, and (c) Low Proficiency Listeners 111
Figure 3.16 Correlation between L1 Spanish Consistency and L2 English Consistency 112
Figure 3.17 Proportion of Fixations on /b/-Onset Pictures and /p/-Onset Pictures as a Function of VOT and Gradiency in (a) Spanish and (b) English 115
Figure 3.18 (a,b) Proportion of Fixations on /b/-Onset Pictures and (c,d) Proportion of Fixations on /p/-Onset Pictures as a Function of VOT and Gradiency for Spanish (Left Column) and English (Right Column) 116
Figure 3.19 (a,b) Recovery Rate and (c,d) Recovery Latency as a Function of tDist and Gradiency for Spanish (Left Column) and English (Right Column) 119
Figure 3.20 Proportion of Fixations on /b/-Onset Pictures and /p/-Onset Pictures as a Function of VOT and Consistency in (a) Spanish and (b) English 121

Figure 3.21 (a,b) Proportion of Fixations on /b/-Onset Pictures and (c,d) Proportion of Fixations on /p/-Onset Pictures as a Function of VOT and Consistency for Spanish (Left Column) and English (Right Column) 123
Figure 3.22 (a,b) Recovery Rate and (c,d) Recovery Latency as a Function of tDist and Consistency for Spanish (Left Column) and English (Right Column) 125
Figure 4.1 Boxplots Showing the Distributions of All Variables included in the Regression Analyses 154
Figure 4.2 VAS Rating across VOT 156
Figure 4.3 Two Representative Participants Showing (a) Low Consistency and High Gradiency, and (b) High Consistency and Low Gradiency in the VAS 156
Figure 4.4 Produced VOTs as a Function of Stimulus VOT in the Imitation Task 157
Figure 4.5 Produced VOT across Stimulus VOT in the Imitation Task 158
Figure 4.6 Two Representative Participants in the Imitation Task: (a) Low Consistency and (b) High Consistency 159
Figure 4.7 Scatterplots Showing the Relationships between VAS Gradiency and (a) Imitation Accuracy, (b) Simple Imitation Accuracy, and (c) Imitation Gradiency, as well as (d) the Correlation between VAS Consistency and Imitation Consistency 161
Figure 4.8 Scatterplots Showing the Relationships between Residualized VAS Gradiency and (a) Imitation Accuracy, (b) Simple Imitation Accuracy, and (c) Residualized Imitation Gradiency 162
Figure 4.9 Scatterplot Showing the Relationship between Auditory Acuity and Imitation Accuracy 163
Figure 4.10 Scatterplot Showing the Relationship between Visual Working Memory and VAS Consistency 164

List of Tables

Table 2.1 Stimulus Sets Used in Experiment 1.1 30
Table 2.2 Number of Trials per VOT Value in the Narrow and Wide Variance Conditions for Spanish and Basque 32
Table 2.3 Stimulus Sets Used in Experiment 1.2 51
Table 2.4 Number of Trials per VOT Value in the Narrow and Wide Variance Conditions for English 52
Table 2.5 Bayesian Evidence for the Distributional Learning Effect 58
Table 3.1 Summary of Tasks and Variables for Experiment 2 80
Table 3.2 Stimuli Used in the English VWP with International Phonetic Alphabet (IPA), as in Kapnoula et al. (2021) 82
Table 3.3 Stimuli Used in the Spanish VWP with IPA and Meanings 83
Table 4.1 Summary of Tasks and Variables for Experiment 3 144
Table 4.2 Descriptive Statistics for All Variables Included in the Hierarchical Regression Models 155
Table 5.1 Summary of Spearman's Correlations between Potential Source Variables and Gradiency and Consistency in L1 Spanish, with BF in Parentheses 179
Table 5.2 Summary of Spearman's Correlations between Unresidualized Gradiency and Consistency, with BF in Parentheses 182

References

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America*, *106*(4), 2031–2039. https://doi.org/10.1121/1.427949

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171–3183. https://doi.org/10.1121/1.1701898

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163–187. https://doi.org/10.1016/0010-0277(94)90042-6

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language, 121,* Article 104279. https://doi.org/10.1016/j.jml.2021.104279

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America, 152*(6), 3728–3745. https://doi.org/10.1121/10.0015201

Audacity Team. (2019). *Audacity (Version 2.3.3) [Computer software]*. https://www.audacityteam.org/

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Baese-Berk, M. (2018). Perceptual learning for native and non-native speech. In K. D. Federmeier & D. G. Watson (Eds.), *The psychology of learning and motivation: Current topics in language* (pp. 1–29). Elsevier Academic Press. https://doi.org/10.1016/bs.plm.2018.08.001

Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech*, *16*(3), 189–195. https://doi.org/10.1177/002383097301600301

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/A:1005653411471

Bejjanki, V. R., Clayards, M., Knill, D. C., & Aslin, R. N. (2011). Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLoS ONE*, *6*(5), e19812. https://doi.org/10.1371/journal.pone.0019812

Benkí, J. R. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In *Proceedings of the 4th International Symposium on Bilingualism* (pp. 240–248). Somerville, MA: Cascadilla Press.

Bicknell, K., Bushong, W., Tanenhaus, M. K., & Jaeger, T. F. (2025). Maintenance of subcategorical information during speech perception: Revisiting misunderstood

limitations. *Journal of Memory and Language*, *140*, 104565. https://doi.org/10.1016/j.jml.2024.104565

Bidelman, G. M., Bernard, F., & Skubic, K. (2025). Hearing in categories and speech perception at the "cocktail party." *PLOS ONE*, *20*(1), e0318600. https://doi.org/10.1371/journal.pone.0318600

Bidelman, G. M., Eisenhut, Z., Borowski, L., Rizzi, R., & Pisoni, D. B. (2026). *Elliptical speech reveals the use of broad phonetic categories aids noise-degraded speech perception.* BioRxiv. https://doi.org/10.64898/2026.01.02.695202

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. https://doi.org/10.1016/j.neuroimage.2013.04.093

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353–1366. https://doi.org/10.1162/0898929054985473

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 13 January 2023 from http://www.praat.org/

Bořil, T., & Skarnitzl, R. (2016). Tools *rPraat* and *mPraat*. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue* (pp. 367–374). Springer International Publishing. https://doi.org/10.1007/978-3-319-45510-5_42

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228. https://doi.org/10.1080/23273798.2017.1325508

Bürkner, P. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software, 100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Bushong, W. (2025). Strong evidence for maintenance of gradient representations during language processing. *Glossa Psycholinguistics*, *4*(1). https://doi.org/10.5070/G6011.40229

Bushong, W., & Jaeger, T. F. (2025). Changes in informativity of sentential context affects its integration with subcategorical information about preceding speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0001443

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, *62*(4), 961–970. https://doi.org/10.1121/1.381590

Carraturo, S., H. Kim, H., Kutlu, E., & McMurray, B. (2025). *Lexical knowledge enhances consistency in speech categorization.* PsyArXiv. https://doi.org/10.31234/osf.io/mdsbh_v1

Caudrelier, T., Ménard, L., Beausoleil, M.-M., Martin, C. D., & Samuel, A. G. (2024). When Jack isn't Jacques: Simultaneous opposite language-specific speech perceptual learning in French–English bilinguals. *PNAS Nexus*, *3*(9), pgae354. https://doi.org/10.1093/pnasnexus/pgae354

Champely, S. (2020). *pwr: Basic functions for power analysis* (R package version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

Chao, K.-Y., & Chen, L.-M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics and Chinese Language Processing, 13*(2), 215–232.

Cheng, B., Dai, X., Xiang, X., Zhang, X., & Zhang, Y. (2025b). *Root mean square error as a robust index of gradient speech perception*. Preprint.org. https://doi.org/10.20944/preprints202510.1806.v1

Cheng, B., Xiang, X., Dai, X., Zou, Y., Zhang, X., & Zhang, Y. (2025a). *Beyond categorical perception: Gradient lexical tone processing revealed by visual analog scale*. Preprint.org. https://doi.org/10.20944/preprints202510.1824.v1

Cheng, H.-S., Niziolek, C. A., Buchwald, A., & McAllister, T. (2021). Examining the relationship between speech perception, production distinctness, and production variability. *Frontiers in Human Neuroscience*, *15*, 660948. https://doi.org/10.3389/fnhum.2021.660948

Chiu, S. L., Toscano, C. M., Toscano, J. C., & McMurray, B. (2025). Listeners rapidly adapt to current conditions: "Good-enough" adaptation in multitalker speech perception. *Journal of Experimental Psychology: General*, *154*(11), 3162–3194. https://doi.org/10.1037/xge0001797

Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and Cognition, 25*(5), 827–840. https://doi.org/10.1017/s1366728922000086

Chládková, K., Podlipský, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 414–427. https://doi.org/10.1037/xhp0000333

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research, 61*(8), 1–20. https://doi.org/10.1044/2018_jslhr-s-17-0392

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*(2), 234–250. https://doi.org/10.1016/0749-596X(91)90005-5

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* [Doctoral dissertation, McGill University]. Dissertation Abstracts International. https://escholarship.mcgill.ca/concern/theses/05741s554

Crinnion, A. M., Heffner, C. C., & Myers, E. B. (2024). Individual differences in the use of top-down versus bottom-up cues to resolve phonetic ambiguity. *Attention, Perception, & Psychophysics*, *86*(5), 1724–1734. https://doi.org/10.3758/s13414-024-02889-4

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology, 8,* Article 522. https://doi.org/10.3389/fpsyg.2017.00522

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351, https://joss.theoj.org/papers/10.21105/joss.05351.

Derawi, H., Reinisch, E., & Gabay, Y. (2022). Increased reliance on top-down information to compensate for reduced bottom-up use of acoustic cues in dyslexia. *Psychonomic Bulletin & Review*, *29*(1), 281–292. https://doi.org/10.3758/s13423-021-01996-9

Desmeules-Trudel, F., & Zamuner, T. S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics*, *81*(5), 1654–1672. https://doi.org/10.3758/s13414-019-01693-9

Desmeules-Trudel, F., & Zamuner, T. S. (2023). Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research*, *39*(2), 333–362. https://doi.org/10.1177/02676583211030604

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset F0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Docherty, G. J. (1992). *The timing of voicing in British English obstruents.* New York, NY: Foris.

Drozdova, P., Hout, R. V., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, *19*(5), 914–920. https://doi.org/10.1017/s136672891600002x

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Eger, N. A., & Reinisch, E. (2019). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition, 41*(1), 179–200. https://doi.org/10.1017/S0272263117000377

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America, 130*(4), EL206–EL212. https://doi.org/10.1121/1.3629144

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Falandays, J. B., Brown-Schmidt, S., & Toscano, J. C. (2020). Long-lasting gradient activation of referents during spoken language processing. *Journal of Memory and Language*, *112*, 104088. https://doi.org/10.1016/j.jml.2020.104088

Farris-Trimble, A., & McMurray, B. (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research, 56,* 1328–1345. http://dx.doi.org/10.1044/1092-4388(2012/12-0145)

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 308–327. https://doi.org/10.1037/a0034353

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi/10.3758/bf03193146

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

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

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. https://doi.org/10.1121/1.396115

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor

control. *The Journal of the Acoustical Society of America*, *142*(4), 2007–2018. https://doi.org/10.1121/1.5006899

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition, 26*(2), 349–356. https://doi.org/10.1017/s027226310426209x

Frye, R. E., Fisher, J. M., Coty, A., Zarella, M., Liederman, J., & Halgren, E. (2007). Linear coding of voice onset time. *Journal of Cognitive Neuroscience, 19*(9), 1476–1487. https://doi.org/10.1162/jocn.2007.19.9.1476

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language, 215,* Article 104919. https://doi.org/10.1016/j.bandl.2021.104919

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(7), 1161–1175. https://doi.org/10.1037/xlm0001213

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, *4*, 340. https://doi.org/10.3389/fpsyg.2013.00340

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/bf03194885

Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(2), e1541. https://doi.org/10.1002/wcs.1541

Gianakas, S. P., & Winn, M. B. (2019). Lexical bias in word recognition by cochlear implant listeners. *The Journal of the Acoustical Society of America*, *146*(5), 3373–3383. https://doi.org/10.1121/1.5132938

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Editions de la Maison des Sciences de l'Homme; Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.001

Giovannone, N., & Theodore, R. M. (2021a). Individual differences in lexical contributions to speech perception. *Journal of Speech, Language, and Hearing Research*, *64*(3), 707–724. https://doi.org/10.1044/2020_jslhr-20-00283

Giovannone, N., & Theodore, R. M. (2021b). Individual differences in the use of acoustic-phonetic versus lexical cues for speech perception. *Frontiers in Communication*, *6*, 691225. https://doi.org/10.3389/fcomm.2021.691225

Giovannone, N., & Theodore, R. M. (2023). Do individual differences in lexical reliance reflect states or traits? *Cognition*, *232*, 105320. https://doi.org/10.1016/j.cognition.2022.105320

Giovannone, N., Kapnoula, E. C., & Toscano, J. C. (2024, November 21–24). *Individual differences in speech perception gradiency in the visual analogue scale task: How stable are they?* [Poster presentation]. The 65th Annual Meeting of the Psychonomic Society, New York, NY, United States.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*(3), 401–424. http://dx.doi.org/10.1016/0022-0965(81)90105-3

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295x.105.2.251

Götz, A., Sensoy, Ö., Schwarzer, G., & Höhle, B. (2025). Distributional learning beyond perceptual reorganization: enhanced discrimination in children and adults after bimodal training. *Language Learning and Development*, 1–20. https://doi.org/10.1080/15475441.2025.2575383

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, *50*(2), 109–125. https://doi.org/10.1016/j.specom.2007.07.003

Gow, D., Jr. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*(1), 133–159. http://dx.doi.org/10.1006/jmla.2000.2764

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, *92*(10), 1–29. https://doi.org/10.18637/jss.v092.i10

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. In *Proceedings of the 16th International Congress of Phonetic Sciences* (Vol. 2, pp. 1893–1896). Saarbrücken, Germany: University of Saarbrücken.

Gür, E., & McMurray, B. (2025). Phonetic variability leads to gradient speech perception. *The Journal of the Acoustical Society of America*, *157*(4_Supplement), A171-A171.

Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.11–10.112). Wiley.

Hammond, R. M. (2001). *The sounds of Spanish: Analysis and application (with special reference to American English)*. Cascadilla Press.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics*, *32*(5), 409–418. https://doi.org/10.3758/bf03202770

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*(1), 65–94. https://doi.org/10.1177/0267658307071601

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Holt, L. L. & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Honda, C. T., Clayards, M., & Baum, S. R. (2024a). Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 50*(4), 370–394. https://doi.org/10.1037/xhp0001191

Honda, C. T., Clayards, M., & Baum, S. R. (2024b). Individual differences in the consistency of neural and behavioural responses to speech sounds. *Brain Research*, *1845*, 149208. https://doi.org/10.1016/j.brainres.2024.149208

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82. https://doi.org/10.3389/fnhum.2011.00082

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

Ibarra Murillo, O. (2013). El code switching vasco-castellano en oraciones atributivas de hablantes jóvenes. *Oihenart*, *28*, 115–130.

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

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49–66.

Jensen, A. R., & Rohwer, W. D. (1966). The stroop color-word test: a review. *Acta Psychologica, 25*(1), 36–93. https://doi.org/10.1016/0001-6918(66)90004-7

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77*(1), 30–60. http://dx.doi.org/10.1006/jecp.1999.2553

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *Big Five Inventory (BFI).* APA PsycTests. https://doi.org/10.1037/t07550-000

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language, 192,* 15–24. https://doi.org/10.1016/J.BANDL.2019.02.004

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language, 223,* Article 105031. https://doi.org/10.1016/j.bandl.2021.105031

Kapnoula, E. C., & Samuel, A. G. (2024). Sensitivity to subphonemic differences in first language predicts vocabulary size in a foreign language. *Language Learning, 74*(4), 950–984. https://doi.org/10.1111/lang.12650

Kapnoula, E. C., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden-paths, but not perception of speech-in-noise. *Journal of Experimental Psychology: Human Perception and Performance, 47*(4), 578–595. https://doi.org/10.1037/xhp0000900

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594–1611. https://doi.org/10.1037/xhp0000410

Kapnoula, E. E. (2016). *Individual differences in speech perception: Sources, functions, and consequences of phoneme categorization gradiency* [Doctoral dissertation, University of Iowa]. https://doi.org/10.17077/etd.feubmitk

Kartushina, N., & Martin, C. D. (2019). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning*, *69*(1), 71–105. https://doi.org/10.1111/lang.12315

Kartushina, N., & Martin, C. D. (2025). Talker-variability in visual feedback articulatory training: Transfer to words, long-term maintenance, and generalization. *The Journal of the Acoustical Society of America*, *157*(3), 2117–2126. https://doi.org/10.1121/10.0036217

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(3), 549–562. https://doi.org/10.1037/xlm0000744

Kempe, V., Bublitz, D., & Brooks, P. J. (2014). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*(2), 349–366. https://doi.org/10.1111/bjop.12092

Kent, R. D., H. Kim, Y., & Chen, L. (2022). Oral and laryngeal diadochokinesis across the life span: A scoping review of methods, reference data, and clinical applications. *Journal of Speech, Language, and Hearing Research*, *65*(2), 574–623. https://doi.org/10.1044/2021_jslhr-21-00396

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, *34*(6), 769–786. https://doi.org/10.1080/23273798.2019.1582787

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics, 81,* Article 100984. https://doi.org/10.1016/j.wocn.2020.100984

Kim, H., Lee, J., & Yang, T.-H. (2025a). *Individual differences in speech categorization and perceptual cue reliance across phonological contrasts*. PsyArXiv. https://osf.io/preprints/psyarxiv/3nmz6_v2

Kim, H., Klein-Packard, J., Sorensen, E., Oleson, J., Tomblin, B., & McMurray, B. (2025b). Speech categorization consistency is associated with language and reading abilities in school-age children: Implications for language and reading disorders. *Cognition*, *263*, 106194. https://doi.org/10.1016/j.cognition.2025.106194

Kim, H., H. Kim, W., McMurray, B., & Yim, D. (2025c). *Speech categorization consistency predicts language and reading abilities in Korean school-age children.* PsyArXiv. https://osf.io/preprints/psyarxiv/g92aw_v3

Kim, H., Tomblin, B., & McMurray, B. (2025d). *Speech categorization consistency predicts general language abilities.* PsyArXiv. https://doi.org/10.31234/osf.io/u46pj_v3

Kim, H., McMurray, B., Sorensen, E., & Oleson, J. (2025e). The consistency of categorization-consistency in speech perception. *Psychonomic Bulletin & Review*, 1–13. https://doi.org/10.3758/s13423-025-02700-x

Kim, H., Muegge, J. B., & McMurray, B. (2025f). *What are you still waiting for? Real-time lexical access requires encapsulated auditory memory.* PsyArXiv. https://doi.org/10.31234/osf.io/anv8b_v1

Kleinschmidt, D. F. (2020). *What constrains distributional learning in adults?* PsyArXiv. https://doi.org/10.31234/osf.io/6yhbe

Kleinschmidt, D. F., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34), 605–610.

Kleinschmidt, D. F., & Jaeger, T. F. (2015a). Inferring listeners' prior beliefs about unfamiliar talkers. https://doi.org/10.13140/RG.2.1.3803.4405

Kleinschmidt, D. F., & Jaeger, T. F. (2015b). 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

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics*, *38*(4), 569–587. https://doi.org/10.1016/j.wocn.2010.08.003

Kong, E. J. (2019). Individual differences in categorical perception: L1 English learners' L2 perception of Korean stops. *Phonetics and Speech Sciences*, *11*(4), 63–70. https://doi.org/10.13064/ksss.2019.11.4.063

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59,* 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J., & Kang, S. (2023). Individual differences in categorical judgment of L2 stops: A link to proficiency and acoustic cue-weighting. *Language and Speech, 66*(2), 354–380. https://doi.org/10.1177/00238309221108647

Kuilin, L., Clayards, M., & Sonderegger, M. (2025). *Assimilation, gradiency and perception-production link in L2: Imitation of Mandarin sibilants by naïve English speakers*. SSRN. https://doi.org/10.2139/ssrn.5518319

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kutlu, E. (2025, April). *Gradiency in speech categorization and its relation to speech intelligibility* [Conference presentation abstract]. *The Journal of the Acoustical Society of America, 157*(Suppl. 1), A169. https://doi.org/10.1121/10.0037789

Kutlu, E., Baxelbaum, K., Sorensen, E., Oleson, J., & McMurray, B. (2024). Linguistic diversity shapes flexible speech perception in school age children. *Scientific Reports, 14*(1), 28825. https://doi.org/10.1038/s41598-024-80430-1

Kutlu, E., Chiu, S., & McMurray, B. (2022). Moving away from deficiency models: Gradiency in bilingual speech categorization. *Frontiers in Psychology*, *13*, 1033825. https://doi.org/10.3389/fpsyg.2022.1033825

Kutlu, E., Kim, H., & McMurray, B. (2026). Longitudinal changes in the structure of speech categorization across school age years: Children become more gradient and more consistent. *Developmental Science*, *29*(1). https://doi.org/10.1111/desc.70085

Lantto, H. (2015). Conventionalized code-switching: Entrenched semantic-pragmatic patterns of a bilingual Basque–Spanish speech style. *International Journal of Bilingualism*, *19*(6), 753–768. https://doi.org/10.1177/1367006914552830

Lee, J., & Park, H. (2024). Acoustic cue sensitivity in the perception of native category and their relation to nonnative phonological contrast learning. *Journal of Phonetics*, *104*, 101327. https://doi.org/10.1016/j.wocn.2024.101327

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

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., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. https://doi.org/10.1037/h0044417

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology: General, 61*(5), 379–388. https://doi.org/10.1037/h0049038

Lim, S., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*(7), 1390–1405. https://doi.org/10.1111/j.1551-6709.2011.01192.x

Lindsay, S., Clayards, M., Gennari, S., & Gaskell, M. G. (2022). Plasticity of categories in speech perception and production. *Language, Cognition and Neuroscience*, *37*(6), 707–731. https://doi.org/10.1080/23273798.2021.2018471

Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 34–11. https://doi.org/10.1177/002383098602900102

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1783–1798. https://doi.org/10.1037/xhp0000092

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism, 7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America, 134*(3), 2213–2224. https://doi.org/10.1121/1.4817845

Llompart, M. (2024). Lexically-guided perceptual recalibration from acoustically unambiguous input in second language learners. *Journal of Phonetics*, *107*, 101366. https://doi.org/10.1016/j.wocn.2024.101366

Llompart, M., & Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, *22*(5), 1085–1100. https://doi.org/10.1017/s1366728918000925

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7*(3), 166–174. https://doi.org/10.3758/Bf03197535

López-Zamora, M., Luque, J. L., Álvarez, C. J., & Cobos, P. L. (2012). Individual differences in categorical perception are related to sublexical/phonological processing in reading. *Scientific Studies of Reading, 16*(5), 443–456. http://dx.doi.org/10.1080/10888438.2011.588763

Luthra, S., Crinnion, A. M., Saltzman, D., & Magnuson, J. S. (2024). Do they know it's Christmash? Lexical knowledge directly impacts speech perception. *Cognitive Science*, *48*(5), e13449. https://doi.org/10.1111/cogs.13449

Luthra, S., Luor, A., Tierney, A. T., Dick, F., & Holt, L. L. (2025). Statistical learning dynamically shapes auditory perception. *Npj Science of Learning*, *10*(1), 41. https://doi.org/10.1038/s41539-025-00328-z

Maas, E. (2017). Speech and nonspeech: What are we talking about? *International Journal of Speech-Language Pathology*, *19*(4), 345–359. https://doi.org/10.1080/17549507.2016.1221995

Magnuson, J. S., Crinnion, A. M., Luthra, S., Gaston, P., & Grubb, S. (2024). Contra assertions, feedback improves word recognition: How feedback and lateral inhibition sharpen signals over noise. *Cognition*, *242*, 105661. https://doi.org/10.1016/j.cognition.2023.105661

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, *16*(1), 68–89. https://doi.org/10.1097/00003446-199502000-00006

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/10.1037/0033-295x.101.4.653

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication, 2*(1), 15–35. https://doi.org/10.1016/0167-6393(83)90061-4

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mattys, S. L., & Scharenborg, O. (2014). Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors. *Psychology and Aging, 29*(1), 150–162. https://doi.org/10.1037/a0035387

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65*(2), 145–160. https://doi.org/10.1016/j.jml.2011.04.004

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, *21*(3), 748–754. https://doi.org/10.3758/s13423-013-0544-7

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11–12), 887–899. https://doi.org/10.1016/j.specom.2010.01.005

Mattys, S. L., Seymour, F., Attwood, A. S., & Munafò, M. R. (2013). Effects of acute anxiety induction on speech perception. *Psychological Science, 24*(8), 1606–1608. https://doi.org/10.1177/0956797612474323

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (Vol. 2, pp. 522–533). Cascadilla Press.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (Vol. 1, pp. 480–490). Cascadilla Press.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, *34*(4), 323–330. https://doi.org/10.3758/bf03203044

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

McMurray, B. (2017). *Nonlinear Curvefitting for Psycholinguistics* (Version 46). Retrieved from https://osf.io/4atgv/

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America, 152*(6), 3819–3842. https://doi.org/10.1121/10.0016614

McMurray, B. (2025, March 27). EyelinkAnalysis. https://doi.org/10.17605/OSF.IO/C35TG

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–246. https://doi.org/10.1037/a0022325

McMurray, B., Apfelbaum, K. S., & Tomblin, J. B. (2022). The slow development of real-time processing: Spoken-word recognition as a crucible for new thinking about language acquisition and language disorders. *Current Directions in Psychological Science*, *31*(4), 305–315. https://doi.org/10.1177/09637214221078325

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Baxelbaum, K. S., Colby, S., & Tomblin, J. B. (2023). Understanding language processing in variable populations on their own terms: Towards a functionalist psycholinguistics of individual differences, development, and disorders. *Applied Psycholinguistics*, *44*(4), 565–592. https://doi.org/10.1017/s0142716423000255

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147–164. https://doi.org/10.1016/j.cognition.2017.08.013

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1344–1362. https://doi.org/10.1044/2014_jslhr-l-13-0196

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. https://doi.org/10.1016/s0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*(1–3), 271–285. https://doi.org/10.1016/00100277(94)90031-0

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505–512. https://doi.org/10.3758/bf03208147

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*(1–3), 106–115. https://doi.org/10.1159/000261764

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272–277. https://doi.org/10.1093/oxfordjournals.bmb.a070866

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Morey, R., & Rouder, J. (2024). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.7, https://CRAN.R-project.org/package=BayesFactor.

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2008). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895–903. https://doi.org/10.1111/j.1467-9280.2009.02380.x

Myers, E. B., Olson, H. E., & Scapetis-Tycer, J. (2024a). Individual differences in accent imitation. *Open Mind*, *8*, 1084–1106. https://doi.org/10.1162/opmi_a_00161

Myers, E., Phillips, M., & Skoe, E. (2024b). Individual differences in the perception of phonetic category structure predict speech-in-noise performance. *The Journal of the Acoustical Society of America*, *156*(3), 1707–1719. https://doi.org/10.1121/10.0028583

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, *113*(5), 2850–2860. https://doi.org/10.1121/1.1567280

Nguyen, N., Lancia, L., Huttner, L., Schwartz, J.-L., & Diard, J. (2024). Listeners' convergence towards an artificial agent in a joint phoneme categorization task. *Glossa Psycholinguistics*, *3*(1). https://doi.org/10.5070/g6011165

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125. https://doi.org/10.1016/j.jml.2016.03.005

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295x.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/s0010-0285(03)00006-9

Olmstead, A. J., Viswanathan, N., Aivar, M. P., & Manuel, S. (2013). Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology*, *4*, 475. https://doi.org/10.3389/fpsyg.2013.00475

Ou, J., & Yu, A. C. L. (2022). Neural correlates of individual differences in speech categorisation: evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience*, *37*(3), 269–284. https://doi.org/10.1080/23273798.2021.1980594

Ou, J., Yu, A. C., & Xiang, M. (2021). Individual differences in categorization gradience as predicted by online processing of phonetic cues during spoken word recognition: Evidence from eye movements. *Cognitive Science, 45*(3), Article e12948. https://doi.org/10.1111/cogs.12948

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning additional languages as hierarchical probabilistic inference: Insights from first language processing. *Language Learning*, *66*(4), 900–944. https://doi.org/10.1111/lang.12168

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics, 69,* 1–11. https://doi.org/10.1016/j.wocn.2018.04.001

Pereira, O., Gao, Y. A., & Toscano, J. C. (2018). Perceptual encoding of natural speech sounds revealed by the N1 event-related potential response. *Auditory Perception & Cognition*, *1*(1–2), 112–130. https://doi.org/10.1080/25742442.2018.1545106

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382–407. https://doi.org/10.1016/j.jneuroling.2010.02.011

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, *116*(4), 2338–2344. https://doi.org/10.1121/1.1787524

Perkell, J. S., Lane, H., Ghosh, S. S., Matthies, M. L., Tiede, M. K., Guenther, F. H., & Ménard, L. (2008). Mechanisms of vowel production: Auditory goals and speaker acuity. *Proceedings of the 8th International Seminar on Speech Production*, December 8–12, Strasbourg, 29–32.

Phillips, S., & Clopper, C. G. (2011). Perceived imitation of regional dialects. *Proceedings of Meetings on Acoustics*, *12*(1), 060002. https://doi.org/10.1121/1.4704668

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/s0140525x04000056

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328–333. https://doi.org/10.1121/1.1914506

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290. https://doi.org/10.3758/bf03213946

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants.

Journal of Experimental Psychology: Human Perception and Performance, *8*(2), 297–314. https://doi.org/10.1037/0096-1523.8.2.297

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

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, *48*(2), 270–292. https://doi.org/10.1016/0022-0965(89)90006-4

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75–86. https://doi.org/10.1037/a0027979

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81–110. http://dx.doi.org/10.1037/0033-2909.92.1.81

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language*, *10*, 243–335. https://doi.org/10.1016/b978-0-12-608610-2.50012-1

Rizzi, R., & Bidelman, G. M. (2024). Functional benefits of continuous vs. categorical listening strategies on the neural encoding and perception of noise-degraded speech. *Brain Research*, *1844*, 149166. https://doi.org/10.1016/j.brainres.2024.149166

Rizzi, R., & Bidelman, G. M. (2025). Consistency in phonetic categorization predicts successful speech-in-noise perception. *JASA Express Letters*, *5*(12), 125203. https://doi.org/10.1121/10.0041846

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics*, *7*(4), 393–402. https://doi.org/10.1016/s0095-4470(19)31072-1

Rosner, B. S., López-Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, *28*(2), 217–224. https://doi.org/10.1006/jpho.2000.0113

RStudio Team. (2023). *RStudio: Integrated Development Environment for R*. Boston, MA. Retrieved from http://www.rstudio.com/

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

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning, 71*(3), 669–715. https://doi.org/10.1111/LANG.12447

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics, 22*(4), 321–330. https://doi.org/10.3758/Bf03199697

Samuel, A. G. (1979). *Speech is specialized, not special* [Doctoral dissertation, University of Illinois at Urbana–Champaign]

Samuel, A. G., & Frost, R. (2015). Lexical support for phonetic perception during nonnative spoken word recognition. *Psychonomic Bulletin & Review*, *22*(6), 1746–1752. https://doi.org/10.3758/s13423-015-0847-y

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Sarrett, M. E., McMurray, B., & Kapnoula, E. C. (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain and Language*, *211*, 104875. https://doi.org/10.1016/j.bandl.2020.104875

Schertz, J., & Paquette-Smith, M. (2023). Convergence to shortened and lengthened voice onset time in an imitation task. *JASA Express Letters*, *3*(2), 025201. https://doi.org/10.1121/10.0017066

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. https://doi.org/10.1016/s0167-6393(02)00094-8

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, *92*(4), 1841–1855. https://doi.org/10.1121/1.403841

Schreiber, E., Onishi, K., & Clayards, M. (2013). Manipulating phonological boundaries using distributional cues. *Proceedings of Meetings on Acoustics, 19*(1), 060298. https://doi.org/10.1121/1.4801082

Schwartz, R. G., Scheffler, F. L. V., & Lopez, K. (2013). Speech perception and lexical effects in specific language impairment. *Clinical Linguistics & Phonetics*, *27*(5), 339–354. https://doi.org/10.3109/02699206.2013.763386

Segal-Feldman, Y., Hitczenko, K., Goldrick, M., Buchwald, A., Roberts, A., & Keshet, J. (2025). Enhancing analysis of diadochokinetic speech using deep neural networks. *Computer Speech & Language*, *90*, 101715. https://doi.org/10.1016/j.csl.2024.101715

Segal, Y., Hitczenko, K., Goldrick, M., Buchwald, A., Roberts, A., & Keshet, J. (2022). DDKtor: Automatic Diadochokinetic Speech Analysis. *Proceedings of Interspeech 2022*, 4611–4615. https://doi.org/10.21437/Interspeech.2022-311

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research, 44*(2), 384–399. http://dx.doi.org/10.1044/1092-4388(2001/032)

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*(2), B35–B44. http://dx.doi.org/10.1016/j.cognition.2005.03.002

Shao, Y., Saito, K., & Tierney, A. (2023). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly*, *57*(1), 33–63. https://doi.org/10.1002/tesq.3120

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

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

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *The Journal of the Acoustical Society of America*, *108*(6), 3030–3035. https://doi.org/10.1121/1.1320474

Shrem, Y., Goldrick, M., & Keshet, J. (2019). Dr. VOT: Measuring positive and negative voice onset time in the wild. *Proceedings of Interspeech 2019*, 629–633. https://doi.org/10.21437/Interspeech.2019-1735

Simon, J., & Winkler, I. (2018). The role of temporal integration in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(11), 1683–1693. https://doi.org/10.1037/xhp0000564

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211. https://doi.org/10.1037/a0016803

Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning. *Frontiers in Psychology, 6,* Article 682. https://doi.org/10.3389/fpsyg.2015.00682

Smith, F. X., & McMurray, B. (2022). Lexical access changes based on listener needs: Real-time word recognition in continuous speech in cochlear implant users. *Ear and Hearing*, *43*(5), 1487–1501. https://doi.org/10.1097/aud.0000000000001203

Sorensen, E., Oleson, J., Kutlu, E., & McMurray, B. (2024). A Bayesian hierarchical model for the analysis of visual analogue scaling tasks. *Statistical Methods in Medical Research*, 9622802241242320. https://doi.org/10.1177/09622802241242319

Souganidis, C., Molinaro, N., & Stoehr, A. (2024). Bilinguals produce language-specific voice onset time in two true-voicing languages: The case of Basque-Spanish early bilinguals. *Linguistic Approaches to Bilingualism*, *14*(3), 370–399. https://doi.org/10.1075/lab.21081.sou

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*(3), 234–249. https://doi.org/10.1037/h0029078

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62*(2), 435– 448. http://dx.doi.org/10.1121/1.381544

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131–136. https://doi.org/10.1016/j.cognition.2010.10.018

Symons, A., Jasmin, K., & Tierney, A. (2026). Speech perception strategies shift instantly. *Cognition, 266,* 106299. https://doi.org/10.1016/j.cognition.2025.106299

Tan, M., & Jaeger, T. F. (2025). Learning to understand an unfamiliar talker: Testing distributional learning as a model of rapid adaptive speech perception. *Cognition*, *265*, 106195. https://doi.org/10.1016/j.cognition.2025.106195

Tan, M., Sabetello, R., Savic, I., & Jaeger, T. F. (2022). Listeners' categorisation behaviour correlates with gradient changes in exposure statistics. In A. Botinis (Ed.), *Proceedings of 13th International Conference of Experimental Linguistics (pp.169–172).* ExLing Society. https://doi.org/10.36505/ExLing-2022/13

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/ 10.1126/science.7777863

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992. https://doi.org/10.3758/s13423-018-1551-5

Theodore, R. M., Monto, N. R., & Graham, S. (2020). Individual differences in distributional learning for speech: What's ideal for ideal observers? *Journal of Speech, Language, and Hearing Research*, *63*(1), 1–13. https://doi.org/10.1044/2019_jslhr-s-19-0152

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America*, *138*(2), 1068–1078. https://doi.org/10.1121/1.4927489

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language*, *184*, 32–42. https://doi.org/10.1016/j.bandl.2018.06.006

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization. *Psychological Science, 21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tufo, S. N. D., & Myers, E. B. (2014). Phonemic restoration in developmental dyslexia. *Frontiers in Neuroscience*, *8*, 134. https://doi.org/10.3389/fnins.2014.00134

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, *62*(6), 1297–1311. https://doi.org/10.3758/bf03212131

Wade, L., Lai, W., & Tamminga, M. (2021). The reliability of individual differences in VOT imitation. *Language and Speech*, *64*(3), 576–593. https://doi.org/10.1177/0023830920947769

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*, 546. https://doi.org/10.3389/fpsyg.2015.00546

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*(1), 48–61. http://dx.doi.org/10.1037/h0084150

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, http://ggplot2.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … , Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Williams, L. (1977a). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289–297. https://doi.org/10.3758/bf03199477

Williams, L. (1977b). The voicing contrast in Spanish. *Journal of Phonetics*, *5*(2), 169–184. https://doi.org/10.1016/s0095-4470(19)31127-1

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America, 147*(2), 852–866. https://doi.org/10.1121/10.0000692

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research, 56*(4), 1097–1107. http://dx.doi.org/10.1044/1092-4388(2012/12-0086)

Wühr, P. (2007). A Stroop effect for spatial orientation. *The Journal of General Psychology, 134*(3), 285–294. https://doi.org/10.3200/genp.134.3.285-294

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, *8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

Zhang, W., Clayards, M., & Sonderegger, M. (2025). Imitation of F0 tone contours by Mandarin and English speakers is both categorical and continuous. *Journal of Phonetics*, *113*, 101457. https://doi.org/10.1016/j.wocn.2025.101457

Zhang, W., Clayards, M., & Torreira, F. (2023). Phonological mediation effects in imitation of the Mandarin flat-falling tonal continua. *Journal of Phonetics*, *101*, 101277. https://doi.org/10.1016/j.wocn.2023.101277

Zimmerman, S. A., & Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *The Journal of the Acoustical Society of America*, *30*(2), 152–153. https://doi.org/10.1121/1.1909521

Appendices

Appendix 1: Power Analyses

Appendix 1.1: Experiment 1

An a priori power analysis was conducted in G*Power (version 3.1.9.3; Faul et al., 2007) with an alpha level of 0.05 and power of 0.8 for an independent-samples t-test. The effect size (Cohen's $d = 0.857$) was derived from the t-value ($t(22) = 2.10$) reported for the main effect in the eye-tracking data of Clayards et al. (2008). Based on this analysis, we determined that 18 participants per group would provide sufficient power to detect the expected effect size. Given the four experimental conditions (e.g., Spanish-narrow/Basque-wide, Spanish-wide/Basque-narrow, Spanish-narrow/Basque-narrow, and Spanish-wide/Basque-wide), the total sample size was set to 72 participants.

Appendix 1.2: Experiment 2

We conducted a power analysis with the mixedpower package (Kumle et al., 2021) to estimate the sample size required for this experiment based on the primary finding in Kapnoula et al. (2021): the interaction between distance from target (tDist) and slope (the inverse of gradiency) on the likelihood of recovery. This package treats estimates in the linear mixed effects models, the analysis method used for the current study, as effect sizes, and simulates the power in each sample size based on a dataset. We followed the procedure as described in the supplementary information of Kumle et al. (2021; https://lkumle.github.io/power_notebooks/Scenario1_notebook.html). First, we ran the same model with the same data as in Kapnoula et al. (2021): *Recovery rate ~ tDist * Slope * Target voicing + (tDist|subject) + (tDist|item)*. Then, we set the critical value (t/z value) as 2 because this reflects an alpha level of 5% (Baayen et al., 2008). We ran 1,000 simulations as recommended by Kumle et al. (2021). Based on our results and the fact that 80% is considered a minimum value for power in cognitive sciences, we decided to recruit 70 participants and the estimated power is 84.8%.

Appendix 1.3: Experiment 3

We conducted an a priori power analysis for a hierarchical multiple regression using the pwr package (version 1.3-0; Champely, 2020) in R to determine the minimum sample size required for the study. The model included two blocks of predictors: the first block contained seven variables (perceived difficulty of the imitation task, music perception, articulation ability, openness to experience, attention switching, auditory acuity, and working memory), and the second block included one variable (gradiency or consistency). We assumed an effect size of $f^2 = 0.206$, based on D. Kim and Clayards (2019), who reported a relationship between speech perception and production of vowels in an explicit imitation task. Although their study focused on vowels rather than consonants, it represents the most comparable available evidence. With an alpha level of 0.05 and desired power of 0.95, the analysis indicated that a minimum of 73 participants would be required to detect the effect of the main predictor (gradiency or consistency) in the hierarchical regression. To be conservative, 80 participants were recruited.

Appendix 2: Tables for Experiment 1

Table A2.1

Participant Characteristics for Experiment 1.1 (N = 72)

Measure	Min	Max	*M*	*SD*
Age[1]	18	35	25.8	4.7
Spanish AoA[1]	0	8	1.1	1.7
Basque AoA[1]	0	3	0.7	1.1
Spanish LexTale (in percentage)[1]	77.5	100	93.9	4.7
Basque LexTale (in percentage)[1]	64	100	91.0	6.9
Spanish picture naming (out of 65)[1]	61	65	64.5	0.9
Basque picture naming (out of 65)[1]	55	65	61.5	3.2
Musical training duration (in months)	0	240	49.4	59.2
Language switching composite[2]	0	10	5.5	2.5
Self-rated Spanish listening proficiency (lowest: 1; highest: 9)	7	9	8.8	0.5
Self-rated Basque listening proficiency (lowest: 1; highest: 9)	6	9	8.6	0.6
Spanish listening exposure percentage	20	85	50.3	15.5
Basque listening exposure percentage	5	80	38.5	17.1

Note. AoA = age of acquisition.

[1] Available in the Basque Center on Cognition, Brain and Language (BCBL) database since participants provided this information during registration.

[2] Computed as *(percentage of time speaking with Spanish−Basque bilinguals × self-reported frequency of switching between Spanish and Basque [0−10]) / 100*.

*Categorization Slope ~ Distributional variance * Language + (1 + Language | Subject) + (1 | Item)*

Table A2.2
Categorization Slope Results with Distributional Variance and Language as Predictors for the SNBN and SWBW Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.14	0.17	[−0.18, 0.49]	—
Distributional variance (wide)	−0.45	0.16	[−0.77, −0.13]	3.92
Language (Basque)	0.00	0.21	[−0.43, 0.43]	0.66
Distributional variance × language	0.34	0.21	[−0.06, 0.76]	2.15

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Categorization Slope ~ Distributional variance * Language + (1 + Language + Distributional variance | Subject) + (1 + Distributional variance | Item)*

Table A2.3
Categorization Slope Results with Distributional Variance and Language as Predictors for the SNBW and SWBN Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.17	0.14	[−0.10, 0.46]	—
Distributional variance (wide)	−0.28	0.22	[−0.69, 0.17]	1.41
Language (Basque)	−0.01	0.18	[−0.37, 0.35]	0.48
Distributional variance × language	−0.01	0.28	[−0.57, 0.54]	0.64

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation to competitor ~ Distributional variance * Language + (1 + Language | Subject) + (1 + Distributional variance | Item)*

Table A2.4
Fixation-to-Competitor Results with Distributional Variance and Language as Predictors for the SNBN and SWBW Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−3.51	0.14	[−3.76, −3.23]	—
Distributional variance (wide)	0.40	0.14	[0.13, 0.66]	149.12
Language (Basque)	0.49	0.16	[0.15, 0.79]	23.10
Distributional variance × language	0.12	0.17	[−0.20, 0.45]	0.43

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation to competitor ~ Distributional variance * Language + (1 + Distributional variance + Language | Subject) + (1 + Distributional variance | Item)*

Table A2.5
Fixation-to-Competitor Results with Language and Distributional Variance as Predictors for the SNBW and SWBN Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−3.23	0.12	[−3.44, −2.99]	—
Distributional variance (wide)	0.29	0.14	[0.01, 0.56]	10.78
Language (Basque)	0.37	0.15	[0.07, 0.65]	9.59
Distributional variance × language	0.01	0.20	[−0.38, 0.40]	0.36

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

Table A2.6
Participant Characteristics for Experiment 1.2 (N = 72)

Measure	Min	Max	*M*	*SD*
Age[1,2]	18	35	22.6	4.5
Spanish AoA[1]	0	15	1.2	3.4
English AoA[1]	0	15	3.8	3.2
Spanish LexTale (in percentage)[1]	47.5	100	80.3	17.9
English LexTale (in percentage)[1]	53.8	100	80.9	11.8
Musical training (in months)	0	180	45.6	51.3
Language switching composite[3]	0	9.6	3.0	3.0
Self-rated Spanish listening proficiency (lowest: 1; highest: 9)	5	9	8.4	0.9
Self-rated English listening proficiency (lowest: 1; highest: 9)	5	9	7.9	1.1
Spanish listening exposure percentage	3	90	45.3	23.5
English listening exposure percentage	5	97	46.3	29.6

Note. AoA = age of acquisition.

[1] Available in the Basque Center on Cognition, Brain and Language (BCBL) database since participants provided this information during registration; available from the questionnaire for UConn participants.

[2] One participant declined to report this information ($N = 71$).

[3] Computed as *(percentage of time speaking with Spanish−English bilinguals × self-reported frequency of switching between Spanish and English [0−10]) / 100*.

*Categorization Slope ~ Distributional variance * Language + (1 + Language | Subject) + (1 | Item)*

Table A2.7
Categorization Slope Results with Distributional Variance and Language as Predictors for the SNEN and SWEW Groups

	Estimate	*Est. Error*	*95% CrI*	*BF*
(Intercept)	0.30	0.15	[0.01, 0.58]	—
Distributional variance (wide)	−0.60	0.16	[−0.91, −0.28]	38,662.81
Language (English)	0.14	0.19	[−0.25, 0.51]	0.38
Distributional variance × language	−0.19	0.21	[−0.62, 0.22]	0.92

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Categorization Slope ~ Distributional variance * Language + (1 + Language + Distributional variance | Subject) + (1 | Item)*

Table A2.8
Categorization Slope Results with Distributional Variance and Language as Predictor for the SNEW and SWEN Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.12	0.17	[−0.20, 0.47]	—
Distributional variance (wide)	−0.46	0.16	[−0.77, −0.14]	132.69
Language (English)	0.22	0.21	[−0.22, 0.61]	1.20
Distributional variance × language	0.01	0.21	[−0.41, 0.41]	0.52

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation to competitor ~ Distributional variance * Language + (1 + Language | Subject) + (1 + Distributional variance | Item)*

Table A2.9
Fixation-to-Competitor Results with Distributional Variance and Language as Predictors for the SNEN and SWEW Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−2.47	0.11	[−2.69, −2.25]	—
Distributional variance (wide)	0.00	0.15	[−0.28, 0.29]	0.23
Language (English)	0.03	0.14	[−0.25, 0.31]	0.20

Distributional variance × language	−0.07	0.20	[−0.45, 0.33]	0.40

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation to competitor ~ Distributional variance * Language + (1 + Distributional variance + Language | Subject) + (1 + Distributional variance | Item)*

Table A2.10

Fixation-to-Competitor Results with Distributional Variance and Language as Predictors for the SNEW and SWEN Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−2.49	0.17	[−2.82, −2.17]	—
Distributional variance (wide)	0.16	0.16	[−0.15, 0.47]	2.55
Language (English)	−0.16	0.21	[−0.56, 0.26]	0.38
Distributional variance × language	0.16	0.22	[−0.28, 0.59]	0.61

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Categorization Slope ~ Distributional variance * Experiment + (1 + Distributional variance | Subject) + (1 | Item)*

Table A2.11

Categorization Slope Results with Distributional Variance and Experiment as Predictors for the Same-Variance Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.45	0.09	[0.28, 0.62]	—
Distributional variance (wide)	−0.20	0.12	[−0.43, 0.03]	30,649.45
Experiment (Spanish–English)	−0.34	0.12	[−0.58, −0.10]	5,436.65
Distributional variance × experiment	−0.70	0.16	[−1.02, −0.38]	2,148.70

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Categorization Slope ~ Distributional variance * Experiment + (1 + Distributional variance | Subject) + (1 | Item)*

Table A2.12

Categorization Slope Results with Distributional Variance and Experiment as Predictors for the Mixed-Variance Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.50	0.11	[0.29, 0.73]	—
Distributional variance (wide)	−0.14	0.12	[−0.36, 0.09]	545.57
Experiment (Spanish–English)	−0.59	0.15	[−0.89, −0.32]	Inf
Distributional variance × experiment	−0.49	0.16	[−0.81, −0.18]	60.26

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor. Inf = infinite.

*Fixation to competitor ~ Distributional variance * Experiment + (1| Subject) + (1 + Distributional variance | Item)*

Table A2.13
Fixation-to-Competitor Results with Distributional Variance and Experiment as Predictors for the Same-Variance Groups

	Estimate	*Est. Error*	*95% CrI*	*BF*
(Intercept)	−3.09	0.11	[−3.32, −2.88]	—
Distributional variance (wide)	0.40	0.10	[0.19, 0.60]	3.47
Experiment (Spanish–English)	0.57	0.13	[0.32, 0.83]	109.42
Distributional variance × experiment	−0.39	0.14	[−0.68, −0.11]	10.54

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation to competitor ~ Distributional variance * Experiment + (1 + Distributional variance| Subject) + (1 + Distributional variance | Item)*

Table A2.14
Fixation-to-Competitor Results with Distributional Variance and Experiment as Predictors for the Mixed-Variance Groups

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−3.05	0.11	[−3.27, −2.85]	—
Distributional variance (wide)	0.28	0.09	[0.10, 0.45]	218.38
Experiment (Spanish–English)	0.50	0.13	[0.27, 0.76]	27,020.35
Distributional variance × experiment	0.00	0.13	[−0.25, 0.25]	0.25

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Categorization Slope ~ Distributional variance * Group + (1 | Subject) + (1 + Distributional variance | Item)*

Table A2.15
Categorization Slope Results with Distributional Variance and Group as Predictors

	Estimate	Est. Error	95% *CrI*
(Intercept)	0.31	0.19	[−0.09, 0.70]
Distributional variance (wide)	−0.38	0.23	[−0.78, 0.11]
Group	−0.09	0.23	[−0.54, 0.37]
Distributional variance × group	0.01	0.28	[−0.56, 0.52]

BF
—

Distributional variance for $\text{Subject}_{\text{same}}$ = Inf
Distributional variance for $\text{Subject}_{\text{different}}$ = 32,462.75

Distributional variance × group = 1.00

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor. Inf = infinite.

*Fixation to competitor ~ Distributional variance * Group + (1 | Subject)+ (1 + Distributional variance | Item)*

Table A2.16
Fixation-to-Competitor Results with Distributional Variance and Group as Predictors

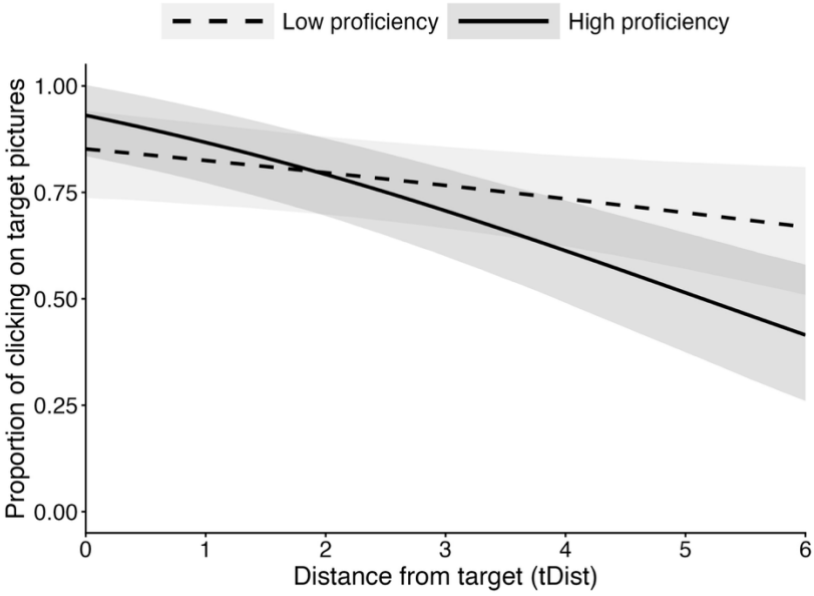
	Estimate	Est. Error	95% *CrI*
(Intercept)	−2.81	0.09	[−2.99, −2.63]
Distributional variance (wide)	0.21	0.08	[0.06, 0.36]
Group	0.02	0.07	[−0.12, 0.15]
Distributional variance × group	0.07	0.09	[−0.11, 0.24]

BF
—
Distributional variance for $Subject_{same}$ = 1.13
Distributional variance for $Subject_{different}$ = 11,483.47
Distributional variance × group = 0.22

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

Appendix 3: Figures for Experiment 2

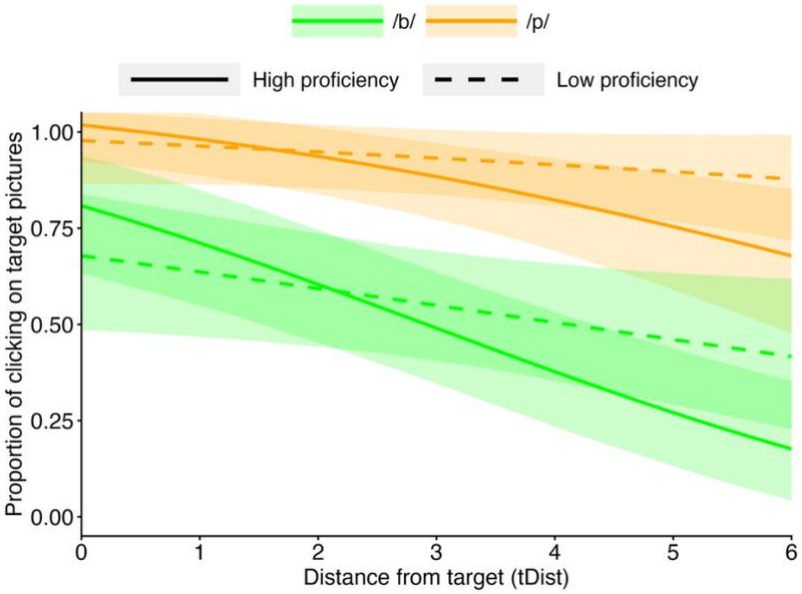
Figure A3.1
The Interaction between tDist and English Proficiency on Predicted Accuracy



Note. "Low proficiency" represents one standard deviation below the mean, and "high proficiency" represents one standard deviation above the mean. These splits are provided for illustrative purposes only, as proficiency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale

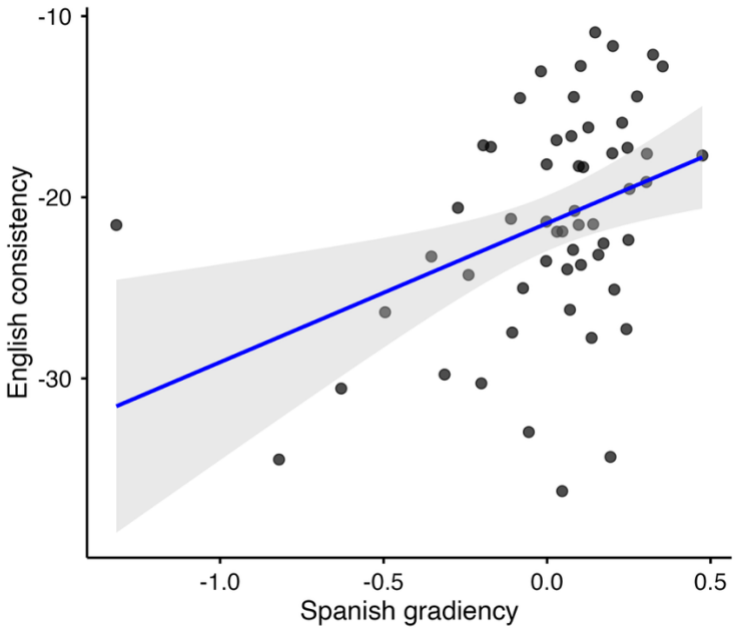
for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

Figure A3.2
Three-way Interaction between tDist, English Proficiency, and Target Voicing on Predicted Accuracy



Note. This three-way interaction is not robust. "Low proficiency" represents one standard deviation below the mean, and "high proficiency" represents one standard deviation above the mean. These splits are provided for illustrative purposes only, as proficiency was modeled as a continuous predictor. Values on both axes were back-transformed from the model's transformed scale to their original scale for better interpretability. Shaded ribbon indicates 95% credible interval around the predicted dependent variable.

Figure A3.3
Correlation between L1 Spanish Gradiency and L2 English Consistency



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant. When the leftmost data point was excluded, the overall data pattern remained unchanged, with a stronger correlation: $\rho = .41$, $BF = 118.78$.

Appendix 4: Tables for Experiment 2

Table A3.1
Participant Characteristics (N = 70)

Measure	Min	Max	*M*	*SD*
Age[1]	18	40	27.8	6.8
Spanish AoA[1]	0	6	0.1	0.8
Basque AoA[1]	0	32	3.1	5.0
English AoA[1]	2	13	6.2	2.2
Percentage of time spent listening to English	0	40	11.3	8.9
Musical training (in months)	0	216	37.4	52.9
Duration of living in English-speaking countries (in months)	0	48	3.5	7.6
Working memory [Corsi] (out of 9)	2	9	6.4	1.5
English picture naming (out of 65)	22	65	50.0	10.3
Inhibitory control [Stroop] (in ms)	12	240	85	38

Note. AoA = age of acquisition.

[1] Available in the Basque Center on Cognition, Brain and Language (BCBL) database since participants provided this information during registration.

*Spanish and English: Accuracy ~ tDist * Splice * TargetVoicing + (tDist|subject) + (tDist|item)*

Table A3.2
Bayesian Model Estimates for Accuracy and tDist for Spanish VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.51	0.11	[0.28, 0.73]	—
tDist	−0.54	0.04	[−0.61, −0.46]	Inf
Splice	0.09	0.01	[0.07, 0.11]	Inf
TargetVoicing	−0.03	0.09	[−0.21, 0.16]	0.40
tDist × Splice	0.01	0.01	[0.00, 0.02]	56.97
tDist × TargetVoicing	0.01	0.03	[−0.05, 0.07]	0.53
Splice × TargetVoicing	−0.02	0.01	[−0.04, 0.00]	11.21
tDist × Splice × TargetVoicing	−0.01	0.01	[−0.02, 0.00]	7.68

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

Table A3.3
Bayesian Model Estimates for Accuracy and tDist for English VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.74	0.15	[0.43, 1.04]	—
tDist	−0.20	0.05	[−0.30, −0.11]	1,999.00
Splice	−0.00	0.01	[−0.02, 0.02]	0.08
TargetVoicing	−0.68	0.13	[−0.95, −0.42]	2,665.67
tDist × Splice	0.00	0.01	[−0.01, 0.02]	1.51
tDist × TargetVoicing	−0.03	0.04	[−0.12, 0.05]	1.61
Splice × TargetVoicing	−0.00	0.01	[−0.02, 0.02]	0.04
tDist × Splice × TargetVoicing	0.01	0.01	[−0.01, 0.02]	2.23

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*Spanish and English: RT ~ tDist * Splice * TargetVoicing + (tDist|subject) + (tDist|item)*

Table A3.4
Bayesian Model Estimates for RT and tDist for Spanish VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	7.29	0.03	[7.22, 7.35]	—
tDist	0.03	0.00	[0.02, 0.04]	Inf
Splice	−0.02	0.00	[−0.03, −0.02]	Inf
TargetVoicing	0.02	0.03	[−0.04, 0.08]	1.00
tDist × Splice	0.00	0.00	[−0.00, 0.00]	1.37
tDist × TargetVoicing	0.00	0.00	[−0.01, 0.01]	0.88
Splice × TargetVoicing	−0.00	0.00	[−0.01, 0.00]	0.27
tDist × Splice × TargetVoicing	−0.00	0.00	[−0.00, 0.00]	0.45

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target.

Table A3.5
Bayesian Model Estimates for RT and tDist for English VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	7.39	0.03	[7.33, 7.44]	—
tDist	0.02	0.00	[0.01, 0.03]	332.33
Splice	0.00	0.00	[−0.00, 0.01]	2.68
TargetVoicing	0.04	0.02	[−0.01, 0.09]	11.38
tDist × Splice	0.00	0.00	[−0.00, 0.00]	0.05
tDist × TargetVoicing	0.00	0.00	[−0.01, 0.01]	0.57
Splice × TargetVoicing	0.01	0.00	[0.00, 0.01]	55.33
tDist × Splice × TargetVoicing	0.00	0.00	[−0.00, 0.00]	0.03

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*Spanish and English: Fixation ~ tDist * Splice * TargetVoicing + (tDist|subject) + (tDist|item)*

Table A3.6
Bayesian Model Estimates for Fixation Proportion and tDist for Spanish VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	2.00	0.18	[1.64, 2.34]	—
tDist	−0.18	0.04	[−0.25, −0.10]	3,999.00
Splice	0.16	0.02	[0.12, 0.21]	Inf
TargetVoicing	−0.16	0.16	[−0.48, 0.15]	2.68
tDist × Splice	−0.01	0.01	[−0.04, 0.01]	2.25
tDist × TargetVoicing	−0.02	0.04	[−0.10, 0.05]	0.97
Splice × TargetVoicing	−0.09	0.02	[−0.13, −0.04]	3,999.00
tDist × Splice × TargetVoicing	0.01	0.01	[−0.01, 0.04]	3.49

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target.

Table A3.7
Bayesian Model Estimates for Fixation Proportion and tDist for English VWP

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	1.40	0.17	[1.05, 1.75]	—
tDist	−0.10	0.03	[−0.16, −0.03]	130.15
Splice	−0.02	0.02	[−0.06, 0.02]	3.54
TargetVoicing	−0.23	0.16	[−0.56, 0.09]	6.45
tDist × Splice	0.01	0.01	[−0.01, 0.03]	2.25
tDist × TargetVoicing	−0.01	0.03	[−0.07, 0.06]	0.23
Splice × TargetVoicing	−0.02	0.02	[−0.05, −0.02]	1.38
tDist × Splice × TargetVoicing	0.00	0.01	[−0.01, 0.02]	0.59

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor.

*Fixation ~ tDist * Splice * TargetVoicing * Language + (tDist|subject) + (tDist|item)*

Table A3.8
Fixation Results for Combined Dataset

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	1.63	0.13	[1.38, 1.89]	—
tDist	−0.14	0.02	[−0.19, −0.10]	Inf
Splice	0.07	0.01	[0.04, 0.10]	7,999.00
TargetVoicing	−0.20	0.10	[−0.39, 0.00]	19.51
Language	−0.27	0.10	[−0.47, −0.06]	64.04
tDist × Splice	0.00	0.01	[−0.02, 0.01]	0.12
tDist × TargetVoicing	−0.01	0.02	[−0.06, 0.03]	1.12
Splice × TargetVoicing	−0.05	0.01	[−0.08, −0.02]	3,999.00
tDist × Language	0.05	0.02	[0.01, 0.09]	38.80
Splice × Language	−0.09	0.02	[−0.12, −0.06]	Inf

TargetVoicing × Language	−0.04	0.10	[−0.23, 0.16]	0.42
tDist × Splice × TargetVoicing	0.01	0.01	[−0.01, 0.02]	3.08
tDist × Splice × Language	0.01	0.01	[0.00, 0.03]	5.75
tDist × TargetVoicing × Language	0.01	0.02	[−0.03, 0.05]	0.64
Splice × TargetVoicing × Language	0.04	0.02	[0.01, 0.07]	81.47
tDist × Splice × TargetVoicing × Language	0.00	0.01	[−0.02, 0.01]	0.68

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target.

*Accuracy ~ tDist * engprof * TargetVoicing + (tDist|subject) + (tDist|item)*

Table A3.9

Bayesian Model Estimates for Accuracy and English Proficiency

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.74	0.16	[0.42, 1.05]	—
tDist	−0.20	0.05	[−0.30, −0.11]	1,332.33
engprof	−0.01	0.01	[−0.02, 0.00]	5.17
TargetVoicing	−0.68	0.14	[−0.96, −0.41]	1,599.00
tDist × engprof	−0.01	0.00	[−0.01, −0.01]	Inf
tDist × TargetVoicing	−0.04	0.04	[−0.12, 0.05]	1.67
Engprof × TargetVoicing	0.00	0.00	[−0.00, 0.00]	3.42
tDist × engprof × TargetVoicing	0.00	0.00	[−0.00, 0.00]	15.91

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target; engprof = English proficiency.

Table A3.10

Spearman's Correlations of Spanish Gradiency, English Gradiency, Spanish Consistency, English Consistency, English Proficiency, Covariate Measures, and Demographic Variables, with BF in Parentheses (N = 70, unless otherwise noted)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Spanish gradiency[1]	—	—	—	—	—	—	—	—	—	—	—	—	—
2. English gradiency[1,2]	.18 (0.39)	—	—	—	—	—	—	—	—	—	—	—	—
3. Spanish consistency	.02 (0.15)	.16 (0.50)	—	—	—	—	—	—	—	—	—	—	—
4. English consistency[2]	**.40 (12.86)**	.04 (0.30)	**.48 (27.18)**	—	—	—	—	—	—	—	—	—	—
5. English proficiency	.17 (0.42)	.02 (0.31)	.12 (0.23)	.21 (1.14)	—	—	—	—	—	—	—	—	—
6. English listening exposure	.04 (0.16)	−.00 (0.42)	.10 (0.21)	.07 (0.39)	**.48 (918)**	—	—	—	—	—	—	—	—
7. Inhibitory control	−.25 (1.24)	−.03 (0.30)	−.08 (0.18)	.22 (1.69)	.03 (0.16)	−.03 (0.15)	—	—	—	—	—	—	—
8. Working memory	−.04 (0.16)	.08 (0.30)	.14 (0.30)	−.03 (0.30)	−.06 (0.17)	−.20 (0.55)	−.10 (0.21)	—	—	—	—	—	—
9. Musical training duration	.15 (0.31)	.10 (0.32)	.14 (0.28)	.08 (0.36)	−.01 (0.15)	.06 (0.17)	−.10 (0.21)	.08 (0.18)	—	—	—	—	—
10. Duration of living in English-speaking countries	.07 (0.18)	.11 (0.37)	.04 (0.16)	.21 (0.48)	.29 (2.63)	.04 (0.16)	−.05 (0.16)	.15 (0.30)	−.12 (0.25)	—	—	—	—
11. Age	.11 (0.23)	−.04 (0.34)	−.06 (0.17)	.03 (0.37)	.05 (0.16)	**−.30 (3.72)**	.03 (0.15)	.12 (0.25)	−.05 (0.16)	.14 (0.28)	—	—	—
12. English AoA	−.04 (0.16)	−.09 (0.32)	.01 (0.15)	−.08 (0.63)	−.19 (0.52)	−.11 (0.22)	−.05 (0.16)	.08 (0.19)	−.04 (0.16)	.07 (0.18)	**.39 (37.60)**	—	—
13. Basque AoA	−.10 (0.21)	−.16 (0.84)	.05 (0.16)	.07 (0.35)	−.12 (0.25)	−.27 (1.76)	−.02 (0.15)	−.09 (0.19)	−.18 (0.45)	−.15 (0.32)	.13 (0.26)	.07 (0.18)	—

Note. Correlations with at least moderate evidence, based on the Bayes Factor, are in bold.

[1] Consistency has been taken into account through residualization; [2] $N = 56$.

*Proportion of fixations on /b/-onset pictures ~ VOTstep * Spanish gradiency + (VOTstep|subject) + (VOTstep|item)*

Table A3.11
Bayesian Model Estimates for the Proportion of Fixations on /b/-Onset Pictures and Gradiency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.22	0.13	[−1.48, −0.96]	—
VOTstep	−0.24	0.04	[−0.32, −0.17]	Inf
Gradiency	0.10	0.18	[−0.26, 0.47]	0.79
VOTstep × Gradiency	−0.02	0.04	[−0.10, 0.07]	0.39

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; VOT = voice onset time.

*Proportion of fixations on /b/-onset pictures ~ VOTstep * English gradiency + (VOTstep|subject) + (VOTstep|item)*

Table A3.12
Bayesian Model Estimates for the Proportion of Fixations on /b/-Onset Pictures and Gradiency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.36	0.13	[−1.62, −1.11]	—
VOTstep	−0.08	0.03	[−0.14, −0.02]	93.12
Gradiency	−0.04	0.16	[−0.36, 0.27]	0.27
VOTstep × Gradiency	0.02	0.03	[−0.03, 0.08]	1.32

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; VOT = voice onset time.

*Proportion of fixations on /p/-onset pictures ~ VOTstep * Spanish gradiency + (VOTstep|subject) + (VOTstep|item)*

Table A3.13
Bayesian Model Estimates for the Proportion of Fixations on /p/-Onset Pictures and Gradiency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.03	0.17	[−1.35, −0.69]	—
VOTstep	0.26	0.04	[0.19, 0.34]	Inf
Gradiency	0.11	0.18	[−0.24, 0.46]	0.86
VOTstep × Gradiency	0.07	0.05	[−0.02, 0.16]	6.95

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; VOT = voice onset time.

*Proportion of fixations on /p/-onset pictures ~ VOTstep * English gradiency + (VOTstep|subject) + (VOTstep|item)*

Table A3.14
Bayesian Model Estimates for the Proportion of Fixations on /p/-Onset Pictures and Gradiency for English

	Estimate	Est. Error	95% *CrI*	*BF*

(Intercept)	−0.86	0.15	[−1.15, −0.57]	—
VOTstep	0.08	0.03	[0.03, 0.13]	144.45
Gradiency	−0.04	0.22	[−0.48, 0.39]	0.18
VOTstep × Gradiency	0.00	0.03	[−0.06, 0.05]	0.02

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; VOT = voice onset time.

*Spanish recovery rate ~ tDist * Spanish gradiency + (tDist|subject) + (tDist|item)*

Table A3.15
Bayesian Model Estimates for Recovery Rate and Gradiency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.68	0.07	[0.54, 0.83]	—
tDist	−0.18	0.02	[−0.22, −0.13]	Inf
Gradiency	−0.24	0.17	[−0.58, 0.10]	5.33
tDist × Gradiency	−0.11	0.06	[−0.23, 0.01]	13.95

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target.

*English recovery rate ~ tDist * English gradiency + (tDist|subject) + (tDist|item)*

Table A3.16
Bayesian Model Estimates for Recovery Rate and Gradiency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.98	0.16	[0.66, 1.29]	—
tDist	−0.04	0.02	[−0.07, −0.01]	35.70
Gradiency	0.13	0.13	[−0.14, 0.39]	1.91
tDist × Gradiency	0.01	0.03	[−0.05, 0.08]	0.53

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*Spanish recovery latency ~ tDist * Spanish gradiency + (tDist|subject) + (tDist|item)*

Table A3.17
Bayesian Model Estimates for Recovery Latency and Gradiency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	6.02	0.03	[5.96, 6.09]	—
tDist	0.01	0.01	[0.00, 0.03]	31.00
Gradiency	−0.09	0.04	[−0.17, 0.01]	38.22
tDist × Gradiency	−0.03	0.01	[−0.05, −0.00]	20.16

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*English recovery latency ~ tDist * English gradiency + (tDist|subject) + (tDist|item)*

Table A3.18
Bayesian Model Estimates for Recovery Latency and Gradiency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	6.19	0.03	[6.13, 6.25]	—
tDist	0.03	0.01	[0.01, 0.04]	887.89

Gradiency	−0.03	0.04	[−0.11, 0.04]	1.56
tDist × Gradiency	6.19	0.03	[−0.01, 0.03]	1.75

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*Proportion of fixations on /b/-onset pictures ~ VOTstep * Spanish consistency + (VOTstep|subject) + (VOTstep|item)*

Table A3.19
Bayesian Model Estimates for the Proportion of Fixations on /b/-Onset Pictures and Consistency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.22	0.13	[−1.48, −0.96]	—
VOTstep	−0.24	0.04	[−0.31, −0.17]	Inf
Consistency	0.02	0.02	[−0.01, 0.06]	6.14
VOTstep × Consistency	−0.01	0.00	[−0.02, −0.00]	15.63

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; VOT = voice onset time.

*Proportion of fixations on /b/-onset pictures ~ VOTstep * English consistency + (VOTstep|subject) + (VOTstep|item)*

Table A3.20
Bayesian Model Estimates for the Proportion of Fixations on /b/-Onset Pictures and Consistency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.36	0.13	[−1.62, −1.10]	—
VOTstep	−0.08	0.03	[−0.14, −0.02]	77.43
Consistency	0.00	0.01	[−0.02, 0.02]	0.10
VOTstep × Consistency	0.00	0.00	[−0.01, 0.00]	1.32

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; VOT = voice onset time.

*Proportion of fixations on /p/-onset pictures ~ VOTstep * Spanish consistency + (VOTstep|subject) + (VOTstep|item)*

Table A3.21
Bayesian Model Estimates for the Proportion of Fixations on /p/-Onset Pictures and Consistency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−1.02	0.16	[−1.34, −0.70]	—
VOTstep	0.26	0.04	[0.19, 0.33]	Inf
Consistency	0.02	0.02	[−0.01, 0.05]	3.72
VOTstep × Consistency	0.01	0.00	[0.00, 0.01]	4.42

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; VOT = voice onset time.

*Proportion of fixations on /p/-onset pictures ~ VOTstep * English consistency + (VOTstep|subject) + (VOTstep|item)*

Table A3.22
Bayesian Model Estimates for the Proportion of Fixations on /p/-Onset Pictures and Consistency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	−0.86	0.15	[−1.15, −0.57]	—
VOTstep	0.08	0.03	[0.03, 0.14]	172.91
Consistency	0.00	0.01	[−0.03, 0.03]	0.03
VOTstep × Con	0.00	0.00	[0.00, 0.01]	8.21

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; VOT = voice onset time.

*Spanish recovery rate ~ tDist * Spanish consistency + (tDist|subject) + (tDist|item)*

Table A3.23
Bayesian Model Estimates for Recovery Rate and Consistency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.68	0.08	[0.53, 0.83]	−
tDist	−0.18	0.02	[−0.23, −0.13]	Inf
Consistency	0.00	0.02	[−0.03, 0.03]	0.07
tDist × Consistency	0.00	0.01	[−0.01, 0.01]	0.62

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; Inf = infinite; tDist = distance from the target.

*English recovery rate ~ tDist * English consistency + (tDist|subject) + (tDist|item)*

Table A3.24
Bayesian Model Estimates for Recovery Rate and Consistency for English

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	0.97	0.16	[0.65, 1.30]	−
tDist	−0.04	0.02	[−0.08, −0.01]	34.09
Consistency	0.00	0.01	[−0.02, 0.01]	0.35
tDist × Consistency	0.00	0.00	[0.00, 0.00]	0.08

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*Spanish recovery latency ~ tDist * Spanish consistency + (tDist|subject) + (tDist|item)*

Table A3.25
Bayesian Model Estimates for Recovery Latency and Consistency for Spanish

	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	6.02	0.03	[5.96, 6.09]	−
tDist	0.01	0.01	[0.00, 0.03]	26.21
Consistency	0.00	0.00	[−0.01, 0.01]	0.05
tDist × Consistency	0.00	0.00	[0.00, 0.00]	0.10

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

*English recovery latency ~ tDist * English consistency + (tDist|subject) + (tDist|item)*

Table A3.26

Bayesian Model Estimates for Recovery Latency and Consistency for English

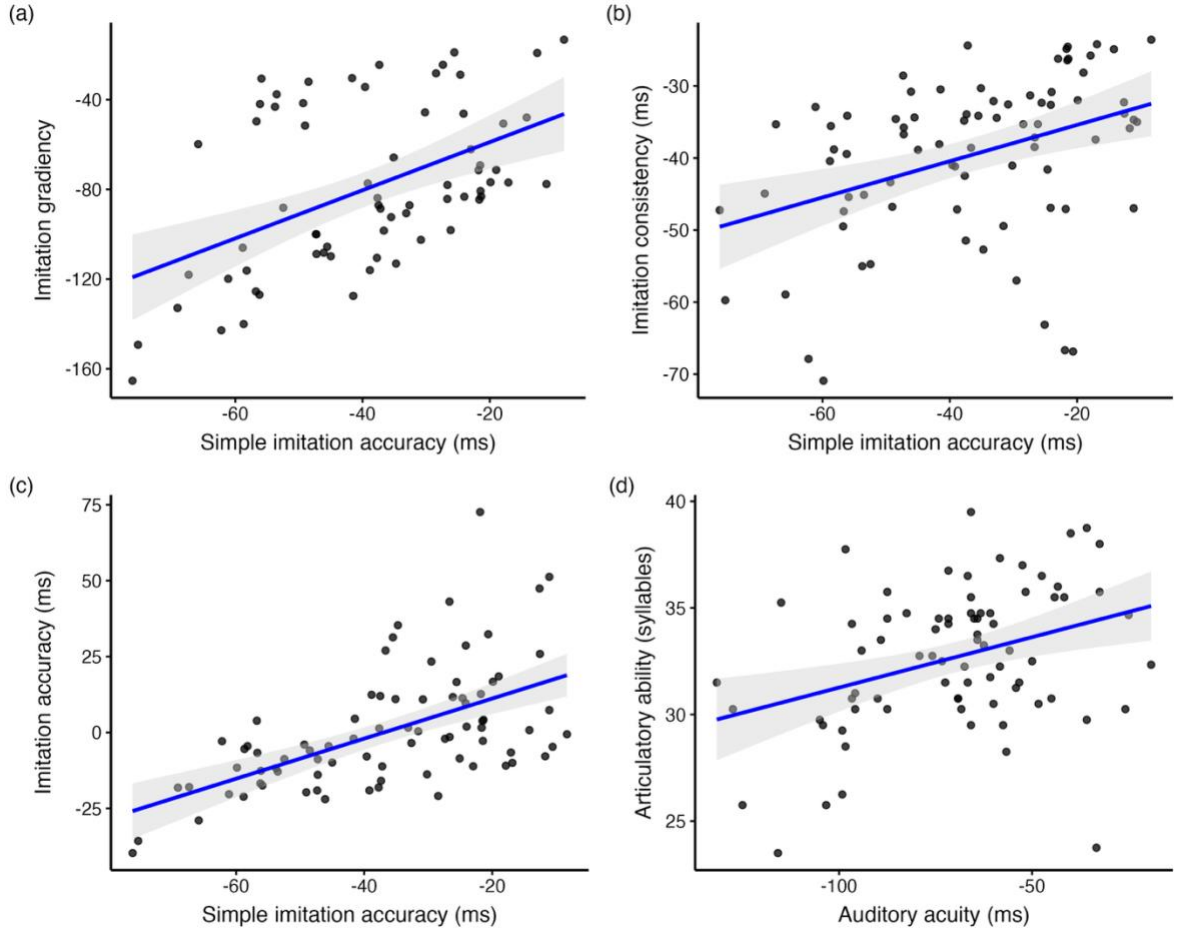
	Estimate	Est. Error	95% *CrI*	*BF*
(Intercept)	6.19	0.03	[6.13, 6.25]	−
tDist	0.03	0.01	[0.01, 0.04]	1,332.33
Consistency	0.00	0.00	[−0.01, 0.00]	0.05
tDist × Consistency	0.00	0.00	[0.00, 0.00]	1.21

Note. Est. error = standard error of the estimate; 95% CrI = 95% credible interval; BF = Bayes Factor; tDist = distance from the target.

Appendix 5: Figure for Experiment 3

Figure A4.1

Scatterplots Showing the Relationships between Simple Imitation Accuracy and (a) Imitation Gradiency, (b) Imitation Consistency, and (c) Imitation Accuracy, as well as (d) the Correlation between Auditory Acuity and Articulatory Ability



Note. The blue line indicates the fitted least-squares regression line, and the grey shaded area represents its 95% confidence interval. Each dot indicates data from one participant.

Appendix 6: Tables for Experiment 3

Table A4.1

Spearman's Correlations among All Main Variables in Experiment 3, with BF in Parentheses (N = 80, unless otherwise noted)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	VAS gradiency	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2.	VAS consistency	−.10 (0.24)	—	—	—	—	—	—	—	—	—	—	—	—	—
3.	Residualized VAS gradiency	**.99 (Inf)**	.00 (0.17)	—	—	—	—	—	—	—	—	—	—	—	—
4.	Imitation gradiency[1]	.16 (0.41)	.06 (0.20)	.16 (0.43)	—	—	—	—	—	—	—	—	—	—	—
5.	Residualized imitation gradiency[1]	.17 (0.48)	−.03 (0.19)	.17 (0.46)	**.95 (∞)**	—	—	—	—	—	—	—	—	—	—
6.	Imitation accuracy	.02 (0.17)	−.06 (0.19)	.01 (0.17)	.14 (0.34)	.12 (0.30)	—	—	—	—	—	—	—	—	—
7.	Simple imitation accuracy	.07 (0.21)	−.10 (0.25)	.07 (0.20)	**.47 (431.28)**	**.33 (7.42)**	**.59 (∞)**	—	—	—	—	—	—	—	—
8.	Imitation consistency	.04 (0.18)	.12 (0.30)	.06 (0.20)	.20 (0.67)	−.06 (0.21)	.06 (0.20)	**.44 (493.88)**	—	—	—	—	—	—	—
9.	Visual working memory	−.17 (0.50)	**.30 (5.79)**	−.13 (0.34)	−.05 (0.20)	−.11 (0.26)	−.05 (0.19)	−.01 (0.17)	.01 (0.17)	—	—	—	—	—	—
10.	Articulatory ability[2]	−.04 (0.18)	.04 (0.19)	−.03 (0.18)	.00 (0.19)	−.01 (0.19)	−.10 (0.24)	.09 (0.23)	.10 (0.25)	.10 (0.25)	—	—	—	—	—
11.	Imitation difficulty	−.08 (0.21)	.05 (0.19)	−.08 (0.22)	−.27 (2.04)	−.26 (1.63)	.08 (0.22)	−.02 (0.17)	−.04 (0.18)	−.14 (0.37)	.14 (0.35)	—	—	—	—
12.	Musical perception skills	−.14 (0.38)	−.03 (0.18)	−.15 (0.40)	.00 (0.19)	−.02 (0.19)	−.06 (0.20)	−.04 (0.18)	.08 (0.22)	.07 (0.21)	.13 (0.31)	−.13 (0.31)	—	—	—
13.	Openness to experience	−.10 (0.25)	−.05 (0.19)	−.11 (0.27)	.00 (0.19)	−.01 (0.19)	.03 (0.18)	.02 (0.17)	.02 (0.17)	−.01 (0.17)	.18 (0.56)	−.01 (0.17)	.24 (1.69)	—	—
14.	Attention switching	−.13 (0.32)	−.10 (0.24)	−.13 (0.33)	−.05 (0.20)	−.09 (0.23)	−.01 (0.17)	.03 (0.18)	.01 (0.17)	.10 (0.25)	−.25 (1.83)	−.12 (0.30)	.26 (2.53)	.11 (0.27)	—
15.	Auditory acuity	.03 (0.18)	.09 (0.23)	.04 (0.18)	.16 (0.44)	.13 (0.33)	.23 (1.23)	.11 (0.27)	.20 (0.83)	.04 (0.18)	**.33 (12.45)**	−.18 (0.59)	.18 (0.59)	.01 (0.17)	.00 (0.17)

Note. Correlations with at least moderate evidence, based on the Bayes Factor, are in bold.

[1] $N = 68$; [2] $N = 77$.

Table A4.2

Hierarchical Regression Predicting Imitation Accuracy

Step	Variable	*β*	R^2	*p*
1	Imitation difficulty	0.15		.201
	Musical ability	−0.10		.429
	Openness to experience	0.11		.446
	Attention switching	−0.10		.473
	Auditory acuity	0.35		.045
	Visual working memory	0.01		.915
	Articulatory ability	−0.24		.102
			.13	
2	VAS gradiency	−0.05		.726
			.14	
			$\Delta R^2 = .00$	.532

Note. Significant result is bolded.

Table A4.3

Hierarchical Regression Predicting Simple Imitation Accuracy

Step	Variable	β	R^2	p
1	Imitation difficulty	−0.02		.861
	Musical ability	−0.10		.421
	Openness to experience	0.12		.373
	Attention switching	0.01		.918
	Auditory acuity	0.13		.357
	Visual working memory	0.02		.847
	Articulatory ability	0.06		.593
			.04	
2	VAS gradiency	0.10		.408
			.05	
			$\Delta R^2 = .01$	.408

Table A4.4

Hierarchical Regression Predicting Imitation Gradiency

Step	Variable	β	R^2	p
1	Imitation difficulty	−0.25		.073
	Musical ability	−0.04		.751
	Openness to experience	0.16		.309
	Attention switching	−0.07		.572
	Auditory acuity	0.17		.246
	Visual working memory	−0.10		.462
	Articulatory ability	−0.01		.927
			.12	
2	VAS gradiency	0.09		.534
			.13	
			$\Delta R^2 = .01$	.191

Table A4.5

Hierarchical Regression Predicting Imitation Consistency

Step	Variable	*β*	R^2	*p*
1	Imitation difficulty	0.01		.946
	Musical ability	0.02		.861
	Openness to experience	−0.01		.946
	Attention switching	0.02		.871
	Auditory acuity	0.16		.203
	Visual working memory	−0.04		.752
	Articulatory ability	0.10		.388
			.05	
2	VAS consistency	0.03		.850
			.05	
			$\Delta R^2 = .00$	.940