ABSTRACT

Title of Dissertation: EVALUATING THE ROLE OF ACOUSTIC CUES IN IDENTIFYING THE PRESENCE OF A CODE-SWITCH

Erika Lynn Exton, Doctor of Philosophy, 2024

Dissertation directed by: Dr. Rochelle S. Newman, Professor, Department of Hearing and Speech Sciences

Code-switching (switching between languages) is a common linguistic behavior in bilingual speech directed to infants and children. In adult-directed speech (ADS), acoustic-phonetic properties of one language may transfer to the other language close to a code-switch point; for example, English stop consonants may be more Spanish-like near a switch. This acoustically-natural code-switching may be easier for bilingual listeners to comprehend than code-switching without these acoustic changes; however, it effectively results in the languages being more phonetically similar at the point of a code-switch, which could make them difficult for an unfamiliar listener to distinguish. The goal of this research was to assess the acoustic-phonetic cues to code-switching available to listeners unfamiliar with the languages by studying the perception and production of these cues.

In Experiment 1 Spanish-English bilingual adults (particularly those who hear code-switching frequently), but not English monolingual adults, were sensitive to natural acoustic cues to code-switching in unfamiliar languages and could use them to identify language switches

between French and Mandarin. Such cues were particularly helpful when they allowed listeners to anticipate an upcoming language switch (Experiment 2). In Experiment 3 monolingual children appeared unable to continually identify which language they were hearing. Experiment 4 provides some preliminary evidence that monolingual infants can identify a switch between French and Mandarin, though without addressing the utility of natural acoustic cues for infants.

The acoustic detail of code-switched speech to infants was investigated to evaluate how acoustic properties of bilingual infant-directed speech (IDS) are impacted by the presence of and proximity to code-switching. Spanish-English bilingual women narrated wordless picture books in IDS and ADS, and the voice onset times of their English voiceless stops were analyzed in code-switching and English-only stories in each register. In ADS only, English voiceless stops that preceded an English-to-Spanish code-switch and were closer to that switch point were produced with more Spanish-like voice onset times than more distant tokens. This effect of distance to Spanish on English VOTs was not true for tokens that followed Spanish in ADS, or in either direction in IDS, suggesting that parents may avoid producing these acoustic cues when speaking to young children.

EVALUATING THE ROLE OF ACOUSTIC CUES IN IDENTIFYING THE PRESENCE OF A CODE-SWITCH

by

Erika Lynn Exton

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Rochelle Newman, Chair
Associate Professor Yi Ting Huang
Associate Professor Jared Novick
Professor Nan Bernstein Ratner
Professor William Idsardi, Dean’s Representative

Acknowledgements

Completing a PhD, while an independent project, is far from a solitary endeavor. If I learned anything from 1.5 years of working fully remotely due to COVID it is that my research is stronger and the process is more fun when completed in the presence of others. This work would not have been possible without the many people who have advised, encouraged, and supported me throughout this process. As I prepare to leave UMD and move to a new university and city, thinking about everybody who made a positive impact on my experience has left me humbled and appreciative of the community I found and built here.

First, thank you to my advisor Dr. Rochelle Newman for six years of consistent mentorship and support. I have grown as a scientist, writer, mentor, and individual through our time together. Thank you for encouraging me to forge my path and coming along for the ride that my research took us both on, especially in the early days when we weren't sure where it would go. Coming to you with new data and getting to talk about interesting or surprising results was always exciting, and helped me to think about my results in new ways. I will miss sitting on the couch in your office and brainstorming new questions, studies, and methods with you. Even after six years, the feedback you provided on my writing was almost always faster and more detailed than I expected; it has turned me into a more confident and careful writer. Thank you for encouraging me to revise my work and keep trying even when initial experiments, manuscript submissions, and grant applications didn't work out and for your investment and belief in me.

Thank you to my parents Erik and Debi, my brother Nick, my sister Kristina, and my grandmother Marilyn Miller for their support throughout my life. They have celebrated my successes at every turn and always believed that there would be more to come. My parents encouraged me to keep asking questions and searching for knowledge from an early age and made it possible for me to pursue my dream of higher-level education and research. This work would not exist without their belief in me, my mom and her SLP buddies introducing me to the concept of thinking about language scientifically, or a friend of my dad's former French co-worker who connected me with my French-Mandarin speaker YM.

Thank you also to the remainder of my dissertation committee, Yi Ting Huang, Jared Novick, Nan Bernstein Ratner, and Bill Idsardi for their feedback on my plan for this work and for discussing early versions of this research with me, for encouraging me to consider the relationship between my research and the broader field, for stimulating scientific conversations, and for flexibility in finding times to meet with me and review my work. Thank you especially to Yi Ting for connecting me with several possible recruitment outlets.

Thank you to the Language Science Center and everybody who has worked there over the years especially Shevaun Lewis, Caitlin Eaves, Tess Wood, Colin Phillips, and Yi Ting Huang. The LSC was a constant source of friendship and camaraderie, scholarship, education, and food. I am grateful for weekly lunches on Thursdays (in all forms), for pastries during morning writing times, and for hot chocolate during Winter Storm. The Language Science Center was one of the key resources that drew me to studying at Maryland, and my experience would have been

vastly different (and almost definitely worse) without the opportunities it provided for me to talk to other language and cognitive scientists at UMD, to learn from their work and perspectives, to improve my presentation skills, and to discover potential new collaborations.

From LDEV, thank you to Amritha Mallikarjun, Christina Blomquist, Maddy Buntrock, Mine Muezzinoglu, Kristin Hoch, Karen Mulak, Katie von Holzen, Meg Cychosz, Jonet Artis, Emily Shroads, Amy Gionfriddo, and Katie Nguyen for your feedback on earlier versions of this research, for asking me difficult questions and encouraging me to think more deeply, for introducing me to new topics, and for your companionship. Amritha and (especially) Emily showed me all the day-to-day ins and outs of doing infant research during my first two years at Maryland. Having Emily as a lab manager who could help trouble-shoot code, think through pivoting to online data collection and recruitment, coordinate a team of undergraduates, and be a friend was an invaluable resource as I began my career as a developmental scientist.

Thank you to all of the students who helped me with aspects of this research, including Ivy Huang for contributing beautiful illustrations, Emily Trujillo for assistance with storyboarding of Spanish wordless picture books, and to Ashley Ruiz, Gabriella Maitra, Richard Saavedra, and Manuela Pietro for assistance with recruitment and data transcription/scoring for the storybook narration study. Thank you also to Erela Imanoel, Lauren Staiger, Morgan Stolker, and Hannah Eisenfeld for countless hours spent hand coding infants' looking behavior and finding minor color changes reflected on infants' faces.

Thank you to my fellow members of the Language Science Outreach team, especially Kathleen Oppenheimer, Lauren Salig, Craig Thorburn, and Alex Krauska. Working as a team with Kathleen, Lauren, and Craig to get our new methods of outreach off the ground provided me with the opportunity to think creatively, gave me a fun project outside my typical research, and reinvigorated my love for sharing language science with others.

Thank you to my final-push writing accountability group, Arynn Byrd and Seongsil Lee, for providing the encouragement I needed to keep making progress at the end. Arynn, I am so proud to see how far we have come, from taking EDMS 641 together our first semester to being the first in our cohort to graduate.

Wednesday evening trivia at Looney's with Allie Johnson, Julianne Garbarino, Zach Maher, Zoe Ovans, Tyler Knowlton, Christina Blomquist, Anna Tinnemore, Kathleen Oppenheimer, and others who passed through helped me to feel welcomed and to build community in my first two (pre-pandemic) years. I appreciated every opportunity to learn from and with my peers, from classes and colloquium talks, to Winter Storm workshops, to casual conversations, to their own research and dissertation defenses.

I was able to begin this PhD directly after finishing my undergraduate degree because of the incredible mentorship I received at Penn State University. My mentors including Janet van Hell, Sarah Grey, Carol Miller, and Chaleece Sandberg taught me how to design and coordinate a research project, built my confidence in my ability to do independent research, and encouraged me to go to grad school. The Center for Language Science at Penn State helped me to envision what an interdisciplinary

approach to language science, where everybody wants to contribute to the success of their colleagues' research, might look like. I doubt that I would have had the confidence to pursue this degree without their mentorship and know that I would not be a bilingualism and code-switching researcher without the formative influence of the Penn State CLS.

Thank you to my current and former roommates: Mira Baum, Emily Brinkman, Mary Kate Byrne, Anna Carmichael, Gillian Hines, Merisa Mitsufuji, and Emily Thomas, for providing me with two years of living in a home with friendship, acceptance, and encouragement. They all believed in my ability to finish this PhD even before they truly knew me and celebrated every small (or big) win they witnessed. Mary Kate took her responsibility to help me break out of my dissertation bubble and think about literally anything else this year very seriously, while always being willing to listen to me talk about data collection, analysis, and interpretation. I am forever grateful for our friendship and for Fridays at Andy's Pizza.

Other social outlets during my time at Maryland, from book clubs and knitting groups, to 9Round Silver Spring, to other friends who stayed in or passed through DC including Caitlin, Rachel, and Megan, brought joy to my life and encouraged me to take the time to explore my non-academic interests and hobbies. Thank you especially to the trainers at 9Round for providing me with a healthy outlet for my frustrations for the past three years.

Throughout my PhD I have been supported by several funding sources. I received a Flagship Fellowship from the Graduate School which partially supported me for the first five years of my degree. I was a Language Science Fellow through NSF

NRT DESE Grant #1449815 and received an NSF Graduate Research Fellowship (thanks in part to mentorship from Shevaun Lewis) through grant DGE 1840340. I also received an NIH F31 pre-doctoral fellowship (F31HD113393) supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development, which funded this dissertation research and the final year of my graduate program. This dissertation does not reflect the official views of the National Science Foundation or the National Institutes of Health.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADS: adult-directed speech

AntEM: anticipatory eye movement [paradigm]

CS: code-switch (code-switching; code-switched)

HASP: high-amplitude sucking paradigm

HPP: headturn preference procedure

IDS: infant-directed speech

L1: first language

L2: second language

VOT: voice onset time

Chapter 1: Introduction: Bilingual language development in the presence of code-switching

Introduction to the general question

Many children are raised in bilingual homes, where they hear multiple languages spoken, sometimes even by the same person in the same conversation. Recent estimates suggest that one third of children under the age of five in the United States hear at least one language in addition to English at home (Migration Policy Institute, 2021). These rates are expected to increase with continued immigration and as young bilingual adults currently living in the United States build families and raise children. In just a few short years, bilingual children go from infants, hearing multiple people speaking to them using multiple languages in a method of exposure that may or may not be systematic in nature and understanding nothing, to young children who communicate effectively with speakers of each of their languages. Along the way, they refine their abilities to discriminate languages using rhythmic difference, discriminate phonetic contrasts, learn words in both languages, and produce speech in a language-specific way. Bilingual children seemingly easily reach a level of language mastery that is only reached with intense effort by people who learn a second language later in life.

The speech perception mechanisms that allow infants to distinguish between their languages and develop separate but intrinsically interactive representations remain unclear. The characteristics of speech that bilingual infants hear do not necessarily, from an adult perspective, make it easy to learn the regularities of each language. Infants sometimes hear code-switching, in which the speaker uses multiple languages in a single sentence or utterance[1].

[1] In this paper I use the term "code-switching" to refer to natural bilingual code-switching; when a bilingual speaker may switch between their two languages in the middle of an utterance (intra-sentential) or between utterances (inter-sentential). Though I am particularly interested in code-

This speech is often produced by bilingual adult speakers who may be accented or less fluent in one of the two languages, and may also be subject to cross-linguistic interference to varying extents. We know that infants can distinguish between languages using rhythmic properties and that infants attend to the distributional regularities of speech sounds in their input in order to develop phonological representations of their language. However, as the majority of child language acquisition research fails to consider multilingualism (Kidd & Garcia, 2022), many of our models of language acquisition do not fully account for how infants learn from input that contains code-switching and other cross-language interactions.

Adult bilingual speakers, particularly when switching between two languages, may produce sounds in one language in ways that are influenced by the other language. Though fluent bilingual adults can use these cues to predict an upcoming code-switch, it is possible that they could cause confusion for listeners less familiar with the languages by blurring the distinction between the languages. As previous studies have focused on the comprehension of a code-switch using semantic information, our current understanding of this process cannot distinguish between language-specific experience and general perceptual sensitivity to such mild phonetic changes. We do not know if these phonetic cues to code-switching are present in infant-directed speech, nor if unfamiliar listeners are able to make use of these fine-grained differences in the distributional phonetics of their speech input to conceptually categorize languages while listening.

The long-term goal of this research is to refine our theories of bilingual language acquisition in order to account for the wide observed variability across infants' language environments, such that we can better understand the perceptual and learning processes that

switching that occurs in a discourse context, previous psycholinguistic research investigating the cognitive processes underlying code-switching has not always used sentences or a discourse. For example, a common paradigm involves asking participants to name words or describe pictures and cuing them with which language to use for each word or trial. When possible these cases of cued switching will be referred to as "language switching" to distinguish them from the "code-switching" that is under the speaker's control and well-attested in corpus studies of bilingual communities.

allow such development to occur. Doing this will improve our ability to provide useful information to parents seeking to raise bilingual children. Though researcher advice has evolved away from earlier messages that bilingualism confuses children and should be avoided, more recent suggestions still lack details about the mechanisms that allow children to learn their two languages and are rarely directly responsive to individual parents' concerns, which could limit the trust that some parents place in our messaging. This work will also contribute to developmental psychology, linguistics, and cognitive science by improving our understanding of the interplay between perceptual discrimination and conceptual categorization in early speech perception.

In this dissertation, I focus on the cues to code-switching available to listeners both in their input and intake. I take a multifaced approach to studying the acoustic properties of code-switching by assessing how listeners *perceive* naturally-occurring acoustic cues to code-switching and how these cues are *produced or provided* in speech to infants. In the perception approach, I study listeners' perceptual sensitivity to acoustic cues to code-switching when categorizing unfamiliar languages and identifying the point of a code-switch. The perception approach includes four experiments, and studies how monolingual and bilingual adults (Exp. 1), monolingual children (Exp. 3) and monolingual infants (Exp. 4) use natural acoustic cues to identify language switches in French and Mandarin, two languages with which they are unfamiliar. Experiment 2 serves as a follow-up to Experiment 1 and examines the time course of these cues. Specifically, it assesses whether adult listeners use cues which occur prior to or after a code-switch in order to detect a switch and identify the correct language. The production approach consists of a single experiment which investigates the acoustic detail of code-switched speech to infants, in order to evaluate how acoustic properties of bilingual infant-directed speech are impacted by the presence of and proximity to code-switching. Spanish-English bilingual women narrated wordless picture books, and I analyzed the voice onset times of their voiceless stops in a code-switching context when speaking to their infant as compared to an adult. Pitch is also assessed in each context, as a cross-check to ensure that mothers used some of the basic properties of infant-directed speech while recording stories to their infants.

In the following sections I will provide an overview of the literature focusing on bilingual language development and particularly the speech that children hear, how they deal with code-switching, and evidence of language differentiation among bilingual children. I will then address other research relevant to code-switching and language discrimination. This dissertation will then move to focus on the perception approach, followed by the production approach. I will then discuss how the results from each of these approaches relate to each other and contribute to our understanding of bilingual speech perception and language development more broadly, including in children with hearing or language disorders, before concluding with future directions for this line of research.

Setting the stage: perceptual influences on language categorization

Take the sentence *"Can you find the gato?"*. A child raised in a Spanish-English bilingual home may regularly hear sentences just like this. In order to comprehend this sentence, as well as its semantic equivalents *"Can you find the cat?"*, *"Puedes encontrar el cat?"*, and *"Puedes encontrar el gato?"*, a child needs to learn several related concepts. They must discover that there are two languages in their environment, which each have their own set of phonemes, sound-meaning mappings, and morphological and syntactic regularities. They also need to learn how these distributions pattern together and which language a given exemplar fits with. Finally, they need to learn the cross-linguistic equivalences. These targets of learning are not necessarily reached in the above order.

Such a situation could be described as a multi-context environment (Qian et al., 2012), as infants must both perceptually detect a change between languages, and adjust some aspect of their processing and learning algorithm in response to that change. If we believe that bilingual infants track distributional regularities of their two languages rather than combining them into a single system, either from birth or after having reached some threshold of experience, then they must be able to react to each context. Infants must also have some representation of each of their language categories, which can serve to store processing algorithms and linguistic detail of each context[2]. As discussed in Byers-Heinlein (2014), conceptual categories of languages are fundamentally different than perceptual discrimination of elements generated by different languages. The conceptual category of a language is composed of many elements and should specify not only which elements comprise the category, but the detail of those elements in terms of both their primary underlying form and how that surface form may vary due to contextual influence. Two elements may have very different surface forms, even if they are generated by the same language (e.g. the phonemes /ɛ/ and /ʒ/ in English). From a perceptual discrimination perspective, these two phones should be distinguished. However, they are both part of the English phonemic inventory, but not a standard Spanish phonemic inventory. /ɛ/ and /ʒ/ should thus be conceptually categorized and considered as a separate conceptual category from /x/ and /ɲ/, which are phones in Spanish but not English. Similar logic can be applied to all linguistic levels, including lexical and morphosyntactic elements. Previous investigations of children's ability to distinguish between languages, including much of the work reviewed below, focus on discrimination of tokens at one linguistic level, rather than assessing the degree to which infants and young children do represent language codes in such conceptual categories that integrate multiple dimensions of linguistic information.

[2] The fact that adult bilinguals do possess representations of language "category" and that those language category representations can influence activation can be shown both by their ability to restrict the lexical items and syntactic structures they use to just one language when necessary (such when speaking to a monolingual) and to produce acoustically distinct realizations of the same phoneme in different language contexts (i.e. to produce a /t/ with a shorter VOT when speaking Spanish than when speaking English) (i.e. Antoniou et al., 2010; Fish et al., 2017; MacLeod & Stoel-Gammon, 2010)

Connectionist models and frameworks including BLINCS (Shook & Marian, 2013) and the PRIMIR framework of the development of speech processing in infancy (Curtin et al., 2011; Werker & Curtin, 2005) can account for some of this learning trajectory by proposing that listeners are attuned to the ways that information across multiple dimensions co-occurs. This distribution of occurrence may influence the contrasts that listeners believe to be important, and thus the formation of categories. For example, we might expect that the existence of categories (at any cognitive level) could influence how individual novel tokens are perceived, such that the relationship between cognitive representation, perception, and input is deeply interactive and continues to change with experience.

The development of these conceptual categories of languages and their relationship to perceptual discrimination is likely to be deeply complex. This complexity may be even greater than is often assumed, including in frameworks that take bilingual language exposure seriously, due to the presence and properties of code-switching in speech to infants. Code-switching could result in speech that is more variable and harder to perceptually discriminate, not only as a function of the languages being less temporally and indexically separated in a code-switching situation than in unilingual speech by bilinguals, but also due to minor ways that the acoustic detail of code-switched speech differs from unilingual speech. This dissertation serves as an exploration of this possibility of added complexity as a result of code-switching.

Rates and types of code-switching in the environment

Multilingual parents employ different approaches regarding how they expose their children to each of the family's languages. The "one-person one-language" method of bilingual language exposure is often discussed as a strategy to aid bilingual language development (Barron-Hauwaert, 2004). It is driven by an underlying assumption that hearing the same person speak multiple languages could cause confusion for their child, and that parents should attempt to minimize such confusion by separating the languages indexically. Such recommendations date back to the early 1900s (Grammont, 1902) and continue today on social media where influencer families document their experience with the OPOL method (mybeautifulblunder, 2023) and bilingual parenting Facebook groups repeat the refrain (Piller & Gerber, 2021). Recent investigations of parental attitudes and experiences with multilingual exposure confirm that parents often make conscious choices about how to expose their children to multiple languages and that they may view OPOL to be 'better' than a mixing strategy (Koelewijn et al., 2023; Piller & Gerber, 2021).

Despite the popularity of this recommendation, there is little experimental evidence directly supporting the benefits of this method as compared to approaches in which caregivers speak all the languages they know. For example, Schott et al. (2023) found no evidence that 5-18 month-old bilingual and monolingual children learn to associate a language with a particular speaker in an experimental setting. Infants did not listen longer to a speaker-language pairing that mismatched an initially trained pairing compared to that initial pairing. This pattern was true even when they saw a picture of the speaker's face while listening, rather than needing to rely solely on acoustic cues to gender and talker identity. This finding suggests that infants do not quickly form speaker-language associations. It is possible that they could do so over a longer period of time with distributed exposures, including through structured language exposure from their families; however, this is an open question as no previous study appears to have investigated whether children growing up in a household following the one-parent one-language link a language to the parent who speaks that language in perception. Proficient adult bilinguals, though, do rapidly form in-lab speaker-language associations to their two known languages (Molnar et al., 2015). Perhaps infants who have lifetime exposure with a one-person one-language pattern at home would similarly be better able to learn a novel speaker-language association in the lab due to a tuning of their expectations.

While children raised in a home following the OPOL method often use both languages in childhood, it is not a necessary approach as children raised in a variety of bilingual language exposure methods successfully acquire multiple languages (Verhagen et al., 2022). Children may actually be less likely to speak a minority language in childhood if it is only spoken at home by one parent, rather than by both parents (de Houwer, 2007; Yamamoto, 2001). Further, even families who intend to take the one-person one-language approach may not follow it strictly (Bail et al., 2015; De Houwer & Bornstein, 2016; Goodz, 1989).

Several measures of the rate of code-switching in children's environments indicate that such behavior is common, though it may vary both among individuals within a community and between communities (Bail et al., 2015; Kremin et al., 2022). In a study of Spanish-English bilingual families living in Maryland involving short play-sessions, Bail et al. (2015) found that all parents code-switched at least once when talking to their 18-24 month-old children. Though the majority of utterances were unilingual, and inter-sentential code-switching was more common than intra-sentential code-switching, overall approximately 15% of utterances or consecutive utterance pairs contained a form of code-switch. For one parent, code-switching occurred between one third of utterances, while other parents produced primarily one language with very little switching.

Day-long recordings with slightly younger French-English families living in Montreal, however, show somewhat lower rates of code-switching (Kremin et al., 2022), though again, all parents of these 10- and 18- month-old infants code-switched at least once during the recording window. Parents either did not change their rate of code-switching or code-switched more often when their children were 18 months old as compared to 10 months old, which may suggest that parents' use of two languages is responsive to what they believe their child knows and understands. Though the structural locations of most parents' code-switching suggested that they primarily used code-switching to improve understanding or provide translation equivalents, other apparent reasonings including emphasis, teaching new vocabulary words, and apparent idiosyncratic borrowings were also common.

Overall, this work shows that code-switching in speech to infants and toddlers is relatively common among bilingual families, despite variability both in the purported methods adopted by families and in their implementation of those methods. Children hear both intra-sentential and inter-sentential code-switching, which may occur for a variety of pedagogical and non-pedagogical reasons.

Outcomes in response to code-switching in the environment

Similarly, research regarding the relationship between children's exposure to code-switching in the home and their vocabulary outcomes is mixed, with evidence for a positive relationship between code-switching and vocabulary, a negative relationship, and no relationship at all. Whereas Place and Hoff (2011, 2016) found that the amount of time during which 2-3 year-old children were exposed to English-only or Spanish-only positively predicted their vocabulary scores in the relevant language, there was no statistically significant relationship between length of time with mixed-language exposure and vocabulary in either language. This finding suggests that time spent hearing both languages produced in a single conversation, including by the same speaker, does not itself harm children's vocabulary development in either language, but also does not provide any benefit to vocabulary. These studies by Place and Hoff (2011, 2016) used a language diary method, in which parents of children acquiring Spanish and English were asked to report which language(s) were spoken to their child during each 30-minute period over the course of seven days, along with the identities of each person speaking to the child and the activities the child engaged in during that block of time. This method allows for a detailed snapshot of how children hear language in their homes, though by summarizing over 30-minute periods of time it does not directly address effects of code-switching within an utterance.

In an analysis of the play-sessions described above (Bail et al., 2015), children whose parents produced more intra-sentential code-switches actually had *larger* vocabularies (both total vocabulary and conceptual) than children whose parents produced less code-switching during the play-session. This finding similarly provides no evidence that code-switching is harmful for vocabulary growth, and in fact may suggest that it is helpful. However, as the directionality of this relationship cannot be assessed, it is possible that the causality is actually in the other direction, with child vocabulary driving parental code-switching. Perhaps parents of children who understood relatively few words chose to code-switch very little, such that child-directed code-switching increases in frequency as child vocabularies increase in size. Such a pattern would be in line with the age effect shown in Kremin et al. (2022), with code-switching exposure increasing from 10- to 18-months of age.

However, as discussed previously, some research also points toward a negative relationship between code-switching exposure and vocabulary. As a different, far less time-intensive approach than either the language diary or the detailed-recording methods, Byers-Heinlein (2013) created a Language Mixing Scale in which parents directly report how often they switched between both languages when talking to their children in a variety of contexts (one-on-one, at home, with friends, with family, at playgroup, and when out). She found a significant negative relationship between parental language mixing and English vocabulary comprehension at 1.5 years in bilinguals who heard a variety of language pairs. Even when accounting for exposure to English, children whose parents reported mixing languages more frequently comprehended fewer English words. It is important to note that both measures in this study were based on parental report and thus could be susceptible to bias; however, parents' estimates of children's relative exposure to each language (Marchman et al., 2017; Orena et al., 2019) and vocabulary (Dale, 1991) have been shown to be relatively reliable with more objective measurements. This Language Mixing Scale also correlated with the amount of mixed-language blocks children heard from the same parent who completed the scale in the sample of 2.5 year olds in South Florida (Place & Hoff, 2016), though not with their mixed-language exposure across all adults.

Further, among 11-month-olds living in France and hearing French and one additional language, the degree to which a given individual only spoke one language to the child was positively correlated with the child's receptive vocabulary in French (Carbajal & Peperkamp, 2020). That is, children whose French-speaking parents focused more on French, without also using the child's additional language, were reported to understand more French words. The same pattern was not true for their other language, or for the degree to which only one language was used during each 30-minute block of time. This finding provides some evidence in support of the one-person-one-language model discussed above. However, these parents were also far more likely to use the OPOL model than other families in France, and the 11-month-olds knew a limited number of words in either language, especially compared to the 18- to 30-month-olds included in similar studies. Other measures of language knowledge (such as volubility as discussed below) may more accurately reflect variability driven by bilingual language exposure method at this age.

Overall, the impact of children's exposure to code-switching in their environment on vocabulary size remains an open question. Several issues face this literature, including the methodological choice in measuring code-switching exposure. Code-switching may impact language development because it means that multiple languages are produced *by the same person*, but it is also possible that the *lack of temporal separation* is in fact a more important contributor. Furthermore, these approaches fail to take into account the structural properties of code-switching and how these properties may impact language development: for example, perhaps code-switching that provides a translation equivalent or re-states an entire phrase in the other language is helpful, but lexical borrowings and switches used to provide emphasis are not. This research also often uses vocabulary size (number of words the child is reported to understand and/or speak) as a measure of language outcome, though this metric is not representative of all aspects of linguistic knowledge and competence even among young children.

In a divergence from the use of vocabulary size, Ruan et al. (2023) found that French-English bilingual infants who heard language-switching more often in general produced fewer vocalizations at 10 months. However, *during mixed-language contexts*, the amount of code-switching that infants heard was positively associated with infant volubility such that they actually vocalized more while they were hearing more code-switching at both 10 and 18 months. This result represents a potentially complex relationship between code-switching, infant engagement, and language development that has not been fully investigated. For example, it could be that hearing code-switching is more engaging due to the use of multiple languages, and that infants therefore vocalize more due to this engagement in the interaction. The effect could also be due to vocabulary knowledge in each language: if these 10- and 18-month-olds understand some words in each language, their overall understanding may be higher during periods of mixed-language input when both languages are in use, even if processing a code-switch invokes a minor processing "cost" (i.e. Morini & Newman, 2019), and they may respond more often due to this understanding. In a single-language situation, though, comprehension *of vocabulary in that language* may be lower than overall vocabulary, leading to slightly lower rates of understanding, engagement, and volubility.

As the studies reviewed above have primarily used relatively global measures of both language exposure and child language ability, we cannot make any conclusions about the function of different types of code-switches on language development. However, the divergence between global and local analyses of switching exposure in Ruan et al. (2023), and between inter-sentential and intra-sentential code-switching in Bail et al. (2015), suggests that a careful examination of code-switching both in the moment and throughout a child's life is important for our understanding of this developmental process.

Effect of code-switching in the moment on processing and word learning

Young bilingual children do appear to process code-switched and unilingual sentences differently (Byers-Heinlein et al., 2017; Kuipers & Thierry, 2012; Morini & Newman, 2019; Potter et al., 2019). That is, when listening to a sentence that switches between two languages, 18-30 month-olds will show increased pupil dilation and spend less time looking at the named object relative to how they behave when listening to a sentence that just uses one language (Byers-Heinlein et al., 2017; Morini & Newman, 2019). However, this patten may be influenced by language dominance, as Spanish-English bilingual toddlers did not show an effect of trial type when listening to a sentence frame in their non-dominant language (Potter et al., 2019). Perhaps when children hear a sentence in their dominant language, they are better able to tune into the details of that language, and expect speech that fits the category to continue. They may have less ability to preferentially activate their non-dominant language when hearing a sentence frame in that language, and therefore process a code-switch into their dominant language just as easily as a unilingual sentence.

It is worth noting that while children look at the correct picture with slightly more accuracy when hearing a unilingual sentence than when hearing a code-switched sentence, they still look at the correct picture significantly above chance in both conditions. Their looking accuracy in the code-switched condition also does not differ from that when listening to a single word without a sentence frame (Morini & Newman, 2019). It is not that hearing a code-switch makes it so difficult for children that they fail to understand. Instead, they may simply need slightly more time to process and understand the word relative to when hearing a unilingual sentence. This effect is not specific to toddlers and young children, as similar processing “costs” in response to code-switching are regularly observed among bilingual adults (Bobb & Wodniecka, 2013; Van Hell & Witteman, 2009 for reviews).

Reports of how bilingual toddlers and young children respond to code-switching in the moment from an electrophysiological standpoint, rather than comprehension, are rare. However, Welsh-English bilingual 2- to 3-year-olds appear to show an earlier ERP response to a language switch than their English monolingual peers (Kuipers & Thierry, 2012). While both groups show a positive ERP when listening to trials that switched into the language which was presented less often (Welsh) than English-only trials that were not preceded by a language switch, they differ in the latency of that effect. Bilinguals show the effect in the 200-300ms time window, and monolinguals in the 300-400ms time window. A similar interaction between participant group and trial type also occurred in a sample of adults (Kuipers & Thierry, 2010). The bilingual adults, however, also showed a later ERP signal associated with re-integration of the unexpected language, which the toddlers did not. As this study did not distinguish between language identity (English vs. Welsh) and experimental status as the expected or unexpected language, or connect the bilingual children’s exposure to or fluency in either language to their performance on the task, it is difficult to determine the exact factors driving this result. The finding does provide some evidence that bilingual children are faster to detect an unexpected language change than monolinguals. This speed is likely driven by their experience with both of the languages in question and particularly their familiarity with phonetic regularities in each language, as the effects occurred prior to the time window which is typically associated with semantic interpretation.

Contrary to previous findings regarding how children process target words which are themselves the code-switched point in a sentence (are produced in a different language than the surrounding sentence frame, or are single words which differ in language from the previous word), children do not appear to show any processing cost or delay in comprehending a word which *follows* an insertional code-switch (Kremin et al., 2023). That is, when hearing a sentence such as "Look! Can you see *el buen* bear?" three-year-old children look just as accurately, if not more accurately, at a picture of a bear as compared to "Look! Can you see the good bear?". While children do seem to notice a language switch on some level, as seen in previous results, this finding suggests that processing "costs" associated with the perception and integration of that switch may be small and resolve quickly, rather than extending through subsequent words. This particular study focused on non-informative adjectives which are produced as an insertional code-switch preceding the target noun, such that the target noun differs in language from the preceding word but is in the same language as the remainder of the sentence frame. In cases of an alternating switch, in which the language only changes once, it is unclear how children process words that follow this switch point.

Code-switching may also interact with children's ability to comprehend speech and remember information, especially in the presence of background noise, and to learn new words. Six- to 11-year-old children process code-switched sentences more slowly than unilingual sentences (Gross et al., 2019). However, accuracy of comprehension questions is similar for code-switched and unilingual sentences or stories among both school-aged children (Gross et al., 2019) and pre-schoolers (Peynircioğlu & Durgunoğlu, 2002). When sentences are presented in the presence of background noise, on the other hand, school-aged children do answer fewer comprehension questions correctly in response to code-switched sentences than unilingual sentences (Gross et al., 2021). Background noise creates a more difficult listening environment such that comprehending speech in such conditions may require listeners to engage more top-down prediction and cognitive control; indeed, children with lower cognitive control skills were worse at answering comprehension questions for code-switched sentences, particularly when those sentences were asked in their non-dominant language.

When learning novel words, Spanish-English bilingual 3-5 year-old children succeed at learning and retaining the meanings of those words when they are presented in a mixed-language context (including natural code-switching), sometimes even above the rate at which they retain word meanings in a single-language context (Kaushanskaya et al., 2023; R. K.-Y. Tsui et al., 2023). However, when a code-switch consisting of a single-word insertion is included in a sentence *prior to* the novel word to be learned (i.e. "Look! Do you see the *perro* on the WALEM?"), three-year-olds do not appear to be able to learn that novel word (Byers-Heinlein et al., 2022). The disparity in these findings yields interesting comparisons: it may be that older children are better able to learn word meanings in the face of code-switching than children under 4. However, from the perspective that the details of the input matter, it is possible that sentence structure and the type and position of a code-switch relative to the novel word influence how children learn from that code-switch. That is, in a longer discourse context and when a code-switch is fully integrated into the sentence in a way that matches existing corpus data (Kaushanskaya et al., 2023), code-switching may be a helpful way to provide the child with useful variability in word meaning and structural contexts. However, single-word insertions may provide variability that the child is not able to meaningfully integrate, especially in a fast-mapping task that probes word-object labels rather than deeper semantic meaning (Byers-Heinlein et al., 2022). More work is needed to further investigate the effect of code-switching on children's word learning. The current work, though, suggests that children do notice when code-switching occurs and that it can influence how they process and integrate both familiar and novel words presented in the same sentence.

Evidence of language separation in children's production

Children's language production allows us a different angle to assess how they think about each of their two languages, including the language(s) they use when talking to different people and how they modulate aspects of phonology for each language. However, it should be noted that studying production in children under the age of five, even those who are typically-developing, is highly limited by their small but growing vocabulary sizes and difficulty meeting articulatory targets for some phonemes. Despite these difficulties, empirical research suggests that bilingual children are sensitive to the language(s) known by their interlocutors and supports a model of growing language differentiation early in childhood, rather than a unitary language system (Genesee, 1989).

In families that primarily follow the one-person one-language method, children show some ability to associate a language with the parent who speaks it by two years of age. They display this ability in production, by using their mother's language more often when speaking to her than when speaking to their father, and vice versa for their father's language (Genesee et al., 1995, 1996). This ability to selectively activate a language in production based on an interlocutor's linguistic knowledge extends to unfamiliar adults, and displays children's ability to generalize these language categories to unknown speakers. Some (though perhaps not all) two-year-olds, when faced with a stranger who appears to be monolingual, speak in ways that indicate some language differentiation and adjustment for this new interlocutor by increasing the proportion of utterances they produce that use only the stranger's language relative to how they speak with their parents (Genesee et al., 1996). All of these abilities appear to be modulated by children's language proficiency and relative dominance in each of their languages, such that they mix languages more often in situations when their non-dominant language is required, relative to when they are able to use their dominant language alone.

Children also notice when a miscommunication occurs due to their adult interlocutor not understanding the language the child used, and will fix these miscommunications by switching to the other language more than by using other methods of repairing a communication breakdown (Comeau et al., 2007). Beyond this global awareness of interlocutor linguistic competence, bilingual two- to three-year-olds also appear to adjust their code-switching behaviors; they will code-switch more often when their bilingual counterpart also code-switches frequently, and produce more unilingual utterances when that same interlocutor, who the children know to be bilingual, uses mostly unilingual utterances (Comeau et al., 2003). It is unclear if the adjustments discussed here are conscious and driven by children's pragmatic inferences within the current interaction, or an automatic pattern-matching behavior to converge with the behaviors of the interlocutor. Regardless of the mechanism, though, adjustment indicates that the children track the properties of the language and rate of code-switching in speech directed to them and have some level of language control, rather than always using best-known lexical items and linguistic structures irrespective of the language which generated that structure.

Sundara et al. (2020) shows that these skills in adapting production to the properties of the environment and speech of an interlocutor may actually appear at an earlier age, before children produce many words. The prosodic contours of 12-month-olds' babbling appears to differ in language-specific ways depending on whether they are talking to an English speaker or a Spanish speaker. Even a small amount of relatively concentrated exposure to Spanish can lead to this more Spanish-like babbling with a Spanish speaking interlocutor, though a complete lack of Spanish exposure prior to meeting the interlocutor was not sufficient for phonetic convergence to occur during the interaction. Both bilingual infants and those who heard a small amount of Spanish in somewhat circumscribed contexts appear to have been able to track some of the phonetic distributional properties of this language separately from their tracking of English, such that they could later reproduce an approximation of those properties in their babbling.

Overall, this work in how children's code-switching differs across conversational contexts and pragmatic demands shows that they do develop early signs of functional differentiation between their languages by two years of age, and that these skills continue to develop through the preschool and early childhood years as their proficiency in each language grows.

Similar patterns are seen in children's phonological development. Different languages often produce the same phoneme in slightly different ways. Assessing productions of such phonemes across languages, then, can give an indication of the extent to which the children differentiate the two languages. That is, if the child produces the same sound in different ways depending on the broader language context, it demonstrates their awareness of language-specific patterns. For bilingual children whose two languages employ different voice onset times for voiced and voiceless phonemes (i.e. English and Spanish, in which English uses positive short-lag VOT[3]s for voiced stops and aspirated, long-lag VOTs for voiceless stops while Spanish uses pre-voicing for voiced stops and positive short-lag VOTs for voiceless stops), their developmental trajectories in producing language-appropriate VOTs can allow us to test the extent to which children differentiate languages phonologically.

Early in childhood, there is little evidence that children produce stop consonants in language-specific ways; in fact, their production seems to be driven more by the ease of reaching articulatory targets and acoustic salience of contrasts. A longitudinal case study of a Spanish-English bilingual 19- to 27-month-old revealed that she produced an adult-like voiced-voiceless distinction in English at 27 months (Deuchar & Clark, 1996). She did not do so in Spanish, though her VOTs in Spanish were shorter than those in English, suggesting some early

[3] Voice onset time, or VOT, is a continuous measurement which can be calculated for stop consonants. It is defined as the length of time between the release of the stop closure and the onset of periodicity associated with voicing. Phonologically voiceless consonants are produced with longer (more positive) VOTs than phonologically voiced consonants. However, languages differ in the range of VOTs used for each type of consonant. Spanish, for example, tends to have shorter VOTs than English. Spanish voiced consonants are often produced with pre-voicing (negative VOT), while voiceless consonants are produced with a short positive lag between the stop release and the onset of voicing (~0-30 ms). Meanwhile, English voiced consonants are typically produced with a short positive lag (~0-30 ms), and voiceless consonants are produced with a much longer positive lag (~30-90 ms). See Lisker & Abramson (1964) for a seminal discussion of cross-linguistic differences in voice onset time.

differentiation. Similarly, German-Spanish[4] bilingual 2- to 3-year-old children differ in the extent to which they display language-specific voicing patterns; some use patterns associated with both target languages, while others seem to employ voicing in a consistent manner regardless of language (Kehoe et al., 2004).

These patterns appear to become more language specific and consistent across individuals as children age. English tends to use longer VOTs than Japanese does, and English-Japanese four-year-olds produce greater VOTs when speaking English than when speaking Japanese, though two-year-olds do not differ between languages (C. E. Johnson & Wilson, 2002). Spanish-English bilingual children differentiate the two languages in their production of voiceless stops at 5- to 6- years of age, while language-specific production in the voiced category may not develop until children are 10 years old (Muru & Lee, 2017). Acquiring these voicing systems in each language may be dependent on degree and context of language exposure, as Dutch-German bilingual 3- to 5-year-old children exposed to more German also produce German with more monolingual-like aspiration (Stoehr et al., 2018).

The evidence reviewed above consists of a small sample of a large literature designed to explore phonological development in bilingual children. Although specific patterns differ among language pairs, aspects of the input and amount of type of exposure, and idiosyncratic differences across children, in general they suggest that while this phonological development is subject to cross-linguistic influence and bilinguals do not necessarily produce sounds the same way as monolinguals, language differentiation at the phonological and phonetic levels does occur in production during the childhood years.

[4] German VOTs are similar to English VOTs

Evidence of language separation in perception of individual phonemes

Bilingual children also perceive speech in a language-specific manner, and appear to do so with increasing precision as they age and their representations of the two languages grow in specificity. Infants in English-Spanish bilingual homes may use the rhythmic distance between the two languages to track them separately, resulting in a greater ability to distinguish a phoneme contrast than their peers learning a rhythmically similar language pair (Sundara & Scutellaro, 2011). This work provides evidence that perceptual discrimination between languages may lead to unique processing approaches for each language. However, such models have not otherwise been tested in infants under one year of age. Further, bilingual infants learning a pair of rhythmically similar languages (Spanish-Catalan) do show some ability to distinguish that same phoneme contrast when tested in a more sensitive method (Albareda-Castellot et al., 2011). The extent to which infants under a year have truly developed conceptual categories of their languages in a way that is robust enough to influence perception of individual phonemes therefore remains unclear.

Older bilingual children do appear to develop separable contrasts for each language in perception, in line with their production of these same phonemes. Spanish-English bilingual preschoolers are more likely to associate a stop phoneme produced using a 0-50ms VOT with the voiceless category when in a Spanish context ("tos" *cough* vs. "dos" *two*) than when they are in an English context ("toe" vs. "doe") (Montanari et al., 2023). Similar results are found in English-speaking school-aged children enrolled in a French immersion program and in Spanish-English bilingual adults. These French-learning children have a greater VOT boundary for the voiced-voiceless contrast when listening to English words as compared to French words (Netelenbos & Li, 2013). Spanish-English bilingual adults identify a nonword with a 0-20ms VOT as beginning with a voiceless stop ("pafri" vs. "bafri") more when other segmental information (the acoustic realization of "fri") leads them to believe the word is in Spanish than when they believe it to be English (J. V. Casillas & Simonet, 2018; Gonzales & Lotto, 2013). Adults further appear to ultimately reach a point at which hearing a voice onset time associated with one language serves as a cue to preferentially activate lexical items in that language (Ju & Luce, 2004). The relationship between the input that children hear in each language and their developing cognitive skills and lexical knowledge that supports these growing language-specific perceptual and phonological systems remains somewhat unclear.

Use of rhythm to perceptually discriminate languages

Perceptual discrimination of language passages, rather than specific phonemes, appears to be affected by rhythmic and durational properties of each passage. Languages have historically been typologically categorized into one of three rhythmic classes: stress-timed, syllable-timed, and mora-timed (Abercrombie, 1967). These categories are not perfect, and indeed more recent theories of linguistic rhythm propose that it exists on a spectrum and that a particular language's position on that spectrum is driven by several phonological properties (i.e. Dauer, 1983, 1987). However, in some ways these categories do correspond to both acoustic properties and human speech perception. Infants, including newborns, prefer listening to their mother's language over rhythmically dissimilar non-native languages and furthermore, can distinguish rhythmically dissimilar languages even if both are unfamiliar (i.e. Byers-Heinlein et al., 2010; Mehler et al., 1988; Moon et al., 1993; Nazzi et al., 1998; Ramus et al., 1999; Ramus, 2002). In some of these cases the speech was low-pass filtered at 400 Hz or resynthesized to remove segmental information (i.e. Ramus, 2002), leading to further evidence that infants use intonation, rhythm, and potentially broad phonotactic information, rather than detailed segmental cues.

However, as noted by Gasparini, et al., (2021), language discrimination using pairs of rhythmically *similar* languages has rarely been tested in newborns. Later in development, at approximately 2-7 months of age, there is evidence that infants can discriminate between languages of both the same and different rhythmic classes (i.e. Bosch & Sebastián-Gallés, 1997; Chong et al., 2018; Christophe et al., 2003; Molnar et al., 2014; Nacar Garcia et al., 2018; Nazzi et al., 2000; Shafer et al., 1999; Zacharaki & Sebastian-Galles, 2021), though conflicting evidence exists (i.e. Chong et al., 2018; Peña et al., 2010). While most studies examining discrimination of rhythmically similar languages suggest that at least one of the languages tested must be infants' native language for them to succeed at discrimination, White et al., (2016) found evidence that 5-month-old English monolingual infants could discriminate Spanish and French, which are both considered to be syllable-timed under classic categorizations. This literature remains unclear regarding the age at which infants become able to discriminate between rhythmically similar non-native languages and the exact properties in the speech signal to which they are attending in order to solve this problem.

Our experience of rhythm in both infancy and adulthood may not be driven by rhythm classes themselves, but rather relative durations of vowel and consonant segments, which differ between rhythm classes due to phonological properties of the languages (Ramus et al., 1999). Duration variability can be calculated in a variety of ways including the proportion of a sentence made up of vocalic intervals (%V) and consonantal intervals (%C; these values are perfectly complementary), the standard deviation of vocalic intervals in a sentence (deltaV; VarcoV) and of consonantal intervals in a sentence (deltaC), the mean of the durational differences between either successive consonant intervals (rPVI-C) or successive vowel intervals (nPVI-V, normalized for speech rate by dividing by their sum), and utterance-final vocalic lengthening (nFinal-V) (Grabe & Low, 2002; Ramus et al., 1999; White et al., 2012; White & Mattys, 2007). These metrics, either independently or in combination, more reliably and accurately predict human language discrimination abilities in infants and adults than rhythm class alone (see Gasparini et al., 2021; Ramus et al., 1999; White et al., 2012, 2014).

Adults, like infants, can distinguish between languages using input that contains durational variability without additional segmental or suprasegmental cues. For example, Ramus et al. (2003) synthesized speech by removing intonational information, replacing all vowels with /a/ and all consonants with /s/ to create “flat sasasa” speech and found that adult French speakers could discriminate English-Spanish and English-Catalan, but not English-Dutch. Further, White, et al., (2012) showed that this effect is due to listeners’ ability to track the difference between durational metrics that exist on a multi-dimensional spectrum rather than rhythmic class membership. English-speaking listeners could distinguish between resynthesized dialects of English which vary with respect to their relative use of final lengthening and interval duration. Similarly, infants can distinguish between two dialects of Catalan (Eastern and Western) and of English (Plymouth and Welsh), which provides further evidence that this discrimination is not based on rhythm “class” *per se* (White et al., 2014; Zacharaki & Sebastian-Galles, 2021). This work indicates that listeners do not require access to semantic information from a speech stream to attend to its acoustic features.

Even when a signal is low-pass filtered to remove detailed segmental information, resynthesized to remove variation in segmental information, or simply in a language that the listener does not understand, listeners track the durations of speech sounds over time. When those relative distributions of durations differ between speech samples, listeners notice. While this perceptual discrimination of languages is not identical to conceptual categorization and may not be sufficient, this literature regarding the role of rhythm indicates that durational features of speech strongly contribute to perceptual discrimination. The applicability of this duration-based perceptual discrimination to mixed-language speech remains to be seen, as listeners may require a certain total duration of speech in each language to compute these statistics over (which is not necessarily available in a dense code-switching context), and rhythmic properties of the two languages may differ in unilingual and mixed language contexts.

Learning of (and discriminating between) multiple novel speech streams simultaneously (statistical learning)

Studies of statistical learning with multiple statistical "languages" offer a unique opportunity to test whether listeners can learn two distinct sets of distributional patterns simultaneously. Such paradigms may create two statistical languages that have unique transitional probabilities, and expose listeners to both languages in an interleaved or sequential manner before testing their ability to segment statistical words according to regularities in one of the languages. Doing so requires listeners to perceptually discriminate the two languages, and to then track a small number of experimentally-specified but unlabeled distributional patterns (transitional probabilities of adjacent or nonadjacent syllables or segments) separately for each of the two statistical languages. These paradigms therefore serve as a simulated bilingual experience.

When adults are provided with an explicit cue to the presence of two statistical languages, such as a pause between the two languages (Gebhart et al., 2009) or a difference in pitch (Weiss et al., 2009), they succeed at tracking distributions over the two languages separately, rather than combining them across structures. However, without such explicit cues, adults appear to either only learn statistical regularities for one of the two languages or to combine distributional statistics across structures. Eight- to 16-month-old monolingual infants show no evidence of learning the properties of multiple speech streams even with a pause or pitch cue to distinguish the languages (Antovich & Graf Estes, 2018; Benitez et al., 2020; Bulgarelli et al., 2017). 14- to 16-month-old bilingual infants, though, can track statistics separately for two structures and succeed at discriminating statistical words from part-words, at least when provided with a short pause or pitch cue between structures during exposure (Antovich & Graf Estes, 2018, 2020). With a linguistic cue to differentiate the speech streams (phonetic; whether the structure was produced in a English-like, French-like, or Finnish-like manner), 9.5- to 12-month-old bilingual *and* monolingual infants succeeded at segmenting words from the second statistical language heard, but not the first (A. S. M. Tsui et al., 2021). While Tsui et al. (2021) did not test bilingual infants over one year of age, who may be able to learn transitional probabilities from multiple structures rather than just one (Antovich & Graf Estes, 2018, 2020), this result shows that monolingual infants have some ability to use phonetic properties of speech to discriminate languages and to use that discrimination to guide perception and learning.

These results suggest that listeners, including infants, can use acoustic cues to distinguish between structures, track statistics over multiple structures, and functionally encapsulate those statistics within each structure. Some type of cue, whether it be phonetic, indexical, or simply a pause between the structures, does appear to be a requisite for this learning, though it is unclear whether such cues are important for the *differentiation* of structures or for the *encapsulation* of distributional statistics within a given structure. Further, monolingual infants may be less able to use such cues than same-aged bilingual infants, perhaps due to a lack of familiarity with the presence of multiple structures in the input more generally. Such a pattern is further supported by Kovács & Mehler (2009b, 2009a) who found differences in monolingual and bilingual 7- and 12- month-old infants' abilities to learn multiple regularities simultaneously or to learn a new regularity that conflicts with a previously-established one. Bilingual 12-month-old infants learned to look to one side of the screen following a nonsense word with an ABA syllable structure, and the other side of the screen following a nonsense word with an ABB syllable structure, while monolinguals only learned one of these structure-side contingencies (Kovács & Mehler, 2009b).

It is worth noting that with the exception of Tsui et al. (2021), the studies reviewed above used non-linguistic cues such as pitch or a short period of silence to distinguish the two statistical languages or syllable structures, which does provide some evidence that linguistic cues are helpful to monolingual infants, unlike a pitch change or pause. None of these experiments have varied durational properties of the two languages, which we know to be a useful cue to discriminate languages for listeners of all ages. Perhaps rhythm, as a property which is operationalized and measured using the durations of phoneme intervals but that incorporates variation over the course of a sentence or passage, has a perceptual saliency which would make it likewise more useful than speaker pitch cues in both this artificial paradigm and in the acquisition of natural languages. Other linguistic cues that may be helpful for listeners and could reflect possible mechanisms for language separation in real-world code-switching environments include phonetic inventories and phonotactic regularities. Testing the use of such cues in an interleaved statistical language learning paradigm is somewhat difficult, however, because it is difficult to assess whether participants are combining statistics across "languages" when the languages consist of non-overlapping segments.

Identifying and discriminating between real languages

A small body of research has considered adults' and children's abilities to identify languages they don't know, or to discriminate between two unfamiliar languages with all segmental and suprasegmental information available. This work shows that adults can identify short samples of a foreign language at above-chance rates and often recognize the geographic region associated with a language, even if they cannot name the language itself (Bond & Fokes, 1991; Lorch & Meara, 1989). Further, adults and school-age children can distinguish between even similar languages at above-chance levels in an ABX task or an AX same-different task (Bond et al., 1998; Lorch & Meara, 1995; Potter & Saffran, 2015; Stockmal et al., 1994), including when both languages are spoken by the same person (Stockmal et al., 1998). Younger children (4-5 years old) correctly indicate when two unfamiliar languages differ at similar rates as older children and adults, which provides further evidence that listeners can attend to the relevant features of languages without the benefit of semantic informativeness across development. However, children and adults may not follow the same patterns with respect to the specific features to which they attend. While adults' performance improved immediately following exposure to one of the to-be-discriminated languages (Bond et al., 1998), children's accuracy was better when familiarized with an *irrelevant* unknown language (Potter & Saffran, 2015). Further, young children struggle to characterize two unique passages of a single unfamiliar language as being "the same" (Stockmal et al., 1994).

The exact properties of the languages that listeners attend to across stages of development and type of previous linguistic experience remains unclear. Performance differs based on the specific language(s) in question: languages which have more distinct phonetic or acoustic properties are more likely to be correctly distinguished. For example, when adult listeners were asked to differentiate between Japanese and one of four other languages which were unidentified for the participants (Arabic, Mandarin, Indonesian, Russian), Arabic and Russian were easily distinguished from Japanese. Meanwhile, Mandarin was reliably confused for Japanese, and the ability to distinguish Indonesian from Japanese varied with respect to participants' language background (Marks et al., 1999, 2001). Bilingual listeners outperformed monolinguals, and German-English bilinguals outperformed Spanish-English bilinguals, suggesting that the specific phonological properties that distinguish two languages *and* how an individual's language experiences influence their perceptual attunement to phonological properties of novel languages influence how well a listener can tell the difference between two languages when listening to speech samples.

However, the operationalization of this necessary relative distinctiveness is missing from these accounts of speech perception. Self-report and multi-dimensional scaling analyses of adult listeners' performance suggests that they use intonation, speech rate, and some language-specific phonemes, in addition to a small number of known lexical items in otherwise unfamiliar languages (Lorch & Meara, 1989; Muthusamy et al., 1994; Stockmal et al., 1996).

Previous research thus provides evidence that adults' and children's sensitivity to rhythmic and phonetic contrasts allows them to use these acoustic and phonetic features to create rudimentary language categories. Some characteristics of these categories may develop with experience over the course of years, as individuals hear individual samples of each language and use those samples to refine their long-term representations, though exposure during the experimental task itself also impacts this discrimination and categorization (Bond et al., 1998). The current project aims to add to this literature by testing how adults and children categorize and distinguish between languages while listening, and crucially the amount of time required to identify a language following a switch.

Processing of code-switching by fluent bilingual adults

As part of the goal of studying language development is to determine the mechanism and trajectory by which children reach later stages of linguistic knowledge, we must also establish how bilingual adults process code-switching and manage representations of multiple linguistic systems. Representations of two languages interact with one another at all levels of processing and production, and are not fully encapsulated systems (Dijkstra & Heuven, 2002; Grosjean & Miller, 1994). For example, lexical and phonological knowledge in a speaker's first language can influence their written and spoken word comprehension in their second language (i.e. Blumenfeld & Marian, 2007; Lagrou et al., 2013; Marian & Spivey, 2003; Spivey & Marian, 1999; Van Assche et al., 2009; Weber & Cutler, 2004). Similar effects also occur in the other language, with a bilingual's L2 and even L3 influencing processing of their L1 (Bice & Kroll, 2015; Midgley et al., 2011; van Hell & Dijkstra, 2002).

However, the broader context does appear to influence the extent of cross-language lexical competition. When listening to a sentence that is highly semantically constraining (i.e. "Before playing, the composer first wiped the keys of the **piano** at the beginning of the concert") cognates are not read any faster than non-cognates, though a cognate facilitation effect remains for sentences which are not semantically constraining (Schwartz & Kroll, 2006). A listener's relative proficiency in each language also influences the strength of this cognate facilitation effect, including the effect of any constraints imposed by the sentence (Lauro & Schwartz, 2017). The language mode (i.e. Grosjean, 1985, 1989, 1997, 1998, 2001) established by the experiment may also influence cross-language competition and interference. When a task appears to only concern one language, some evidence suggests that cross-language lexical competition is weaker than when both languages are used in the experimental context and could be relevant, especially for participants who speak the non-experimental language as an L2 (Canseco-Gonzalez et al., 2010). In the opposite direction, dense exposure to L2 prior to an experiment that takes place entirely in the L2 can decrease the amount of influence of the L1 on L2 processing (Elston-Güttler et al., 2005).

Spoken language production differs from comprehension of written and spoken language in that the speaker knows which language is being used, whereas listeners and readers attempting to interpret a speaker's message may adopt a strategy of considering their entire repertoire of linguistic knowledge in order to maximize chances of comprehension. In cases of picture naming or event description there is evidence of cross-language interaction and competition, though whether this interaction is such that cognates are produced *faster and more accurately* than non-cognates or that competition from the irrelevant language actually *impedes* cognate production is inconsistent and may depend on specific details of language proficiency and task demands (C. D. Martin & Nozari, 2021; Muylle et al., 2022; Starreveld et al., 2014).

In many studies of the processing of code-switching, researchers often discuss a "switch cost" or "processing cost", in which readers or listeners are slower or less accurate to respond to a code-switch, or show a neural correlate associated with additional work required for integration (Bobb & Wodniecka, 2013; Van Hell & Witteman, 2009 for reviews). While this switch cost in unnatural experimental designs consisting of isolated single-word trials is often asymmetric, such that switching into the L1 or more dominant language results in a larger cost than switching into the L2 or less dominant language, these patterns are somewhat less consistent in studies of processing (i.e. reading) language than producing speech. Further, different patterns appear to emerge in more naturalistic situations (Adamou & Shen, 2019; Gosselin & Sabourin, 2021; Van Hell et al., 2015, 2018). For example, Litcofsky & Van Hell (2017) found that when reading sentences with a naturalistic alternating intra-sentential code-switch (as opposed to a single-word insertion, which is less common), Spanish-English bilinguals showed an increased ERP response to switches in the stronger-to-weaker direction, but not the weaker-to-stronger. The stronger-to-weaker direction may require increased engagement in lexical activation and sentence reanalysis processes while both reading and listening to speech (Fernandez et al., 2019; Litcofsky & Van Hell, 2017), while the weaker-to-stronger direction requires a release of inhibition, both in naturalistic sentences and single-word trials (Meuter & Allport, 1999).

Further, recent studies of code-switching have begun to assess potential roles of language mode, linguistic practices of participants' language communities, and the naturalness of code- switching (Wu & Thierry, 2010). Much of this work is driven by the Adaptive Control Hypothesis (Green & Abutalebi, 2013). This hypothesis proposes that the demands of bilinguals' frequent interactional contexts shape how bilinguals control the relative activation of each of their languages; dense code-switching contexts place a higher demand on opportunistic planning while single-language and dual-language contexts place higher demands on goal maintenance and interference control. When the broader context gives participants a reason to expect language switching during the task (i.e. experimenters code-switch; the task is described as using two languages), "switch costs" may be diminished or extinguished entirely (Cheng & Howard, 2008; Gullifer et al., 2013). Further, frequent and infrequent code-switchers differ in the extent to which they demonstrate a "cost" to code-switching, with individuals who engage in code-switching less often for personal idiosyncratic or community norm-driven reasons experiencing a greater code-switching cost (Beatty-Martínez & Dussias, 2017; Gosselin & Sabourin, 2021).

The properties of the speech (or written language) itself can also affect the ease with which bilinguals interpret it. When listening to a Spanish-accented speaker of English who then switches to Spanish, listeners show attenuated N400 effects in comparison to when the speaker does not have a Spanish accent in English (Fernandez, 2015). This effect could be due to the accent serving as a marker of speaker identity which licenses code-switching and makes it more predictable, or the accent itself increasing activation of Spanish before the switch occurs relative to non-accented speech. Listeners find it easier to comprehend code-switches that occur in a common syntactic position in corpora of code-switching than switches that occur in infrequent positions (Beatty-Martínez & Dussias, 2017; Guzzardo Tamargo et al., 2016; Salig et al., 2023). Such patterns have been found using eye-tracking, EEG, and reading times, and in a variety of structural positions of the code-switch including tense (perfect vs. progressive), part of speech (auxiliary vs. participle vs. noun) and determiner-noun (gender-incongruent vs. -congruent). Guzzardo Tamargo et al. (2016) and Beatty-Martínez & Dussias (2017) additionally find that these patterns are true for frequent code-switchers in particular, while ease of comprehension is relatively insensitive to a given code-switching structure's corpus frequency for participants who rarely engage in code-switching. Similarly, Tomić & Valdés Kroff, (2022) show that bilinguals can exploit a relationship between a code-switch and upcoming infrequent information: they anticipate less-frequent words following a code-switch, in line with attested patterns in corpus research.

This work suggests that bilinguals are sensitive to the distribution of the syntactic and lexical properties of code-switching. MacDonald (2013) lays out the Production-Distribution-Comprehension account describing a mechanism by which the ease of language production makes particular structures more common, the distribution of which is apparent to speakers, and then becomes more ingrained as the computations needed to comprehend such structures are used and reinforced more frequently. This process is proposed to be ongoing, such that it continues to be refined over an individual's lifetime and could explain some production errors by children and how they are resolved with increased development and language exposure. However, many developmental consequences of this account are not fully specified.

For example, in the case of the Tomić & Valdés Kroff (2022) study of code-switching and word frequency discussed here, learning the contingent relationship between a code-switch and a low-frequency word requires individuals to track distributions of several dimensions of information: how often they hear a word for a given concept, how often that word appears in each language (English vs. Spanish), how often code-switching occurs in general, and finally the conjunction of code-switching and low-frequency wordforms. For a child early in language acquisition, learning some of these distributions may require a level of detailed linguistic representation that is not yet available, even if the distributions are present in their input and available to be learned. It is possible, then, that children (like adults with less code-switching experience) are not able to use such distribution-based cues to code-switching predictively even if they live in a dense code-switching environment. Though the current work does not assess the developmental trajectory of attunement to these particular content-based distributions in code-switching, it does aim to establish both the presence of a relationship between the distribution of *phonetic* forms and code-switching, and listeners' sensitivity to similar relationships.

Chapter 2: The perception approach: Introduction and stimuli

Introduction

Bilingualism and the presence of phonological systems in multiple languages

Research regarding the interaction between a bilingual's two language in speech perception and production has a long history, aiming to understand the nature of how these phonological inventories are represented (*for a recent special issue focused on this topic see* Amengual, 2021). Bilinguals and second-language learners, particularly those who acquired a language after childhood, are often observed to speak with a "foreign accent" such that phonological properties of one language are applied to speech in the other language (Chang, 2019; Flege, 1995 for examples of reviews). Such patterns are also evident in speech perception, as bilingual listeners may perceive phoneme contrasts and respond to phonetic variation in ways that differ from their monolingual peers and are driven by the specific ways that representations in their two languages interact (Jenkins et al., 1995; López Velarde & Simonet, 2020). Several models have been proposed to explain the detail of this interaction and how it changes over the course of L2 acquisition, including the Speech Learning Model (Flege, 1995), the revised Speech Learning Model (Flege & Bohn, 2021), the Perceptual Assimilation Model of Second Language Learning (Best & Tyler, 2007), the Native Language Magnet Model (Kuhl et al., 2008; Kuhl & Iverson, 1995), and the Second Language Perception Model (Van Leussen & Escudero, 2015).

These models differ in how they propose the interaction between L1 and L2 develops, the relative influence of each language on the other including the nature of influence from the L2 on L1, and the unit at which they propose speech perception occurs. A detailed review of the literature regarding cross-linguistic interaction in bilingualism and second-language speech production and perception, including the linguistic levels at which each model proposes independence or interaction between multiple languages, is outside the scope of this dissertation. However, important findings from this literature do suggest a degree of language-specific allophony, such that at least some speech sounds are produced differently depending on the language context. Development of this allophony is gradual, even in early simultaneous bilingual speakers (Deuchar & Clark, 1996; Kehoe et al., 2004; Muru & Lee, 2017; Stoehr et al., 2018), may not result in phonetic categories that are identical to those used by monolingual speakers in each language, and may depend on the individual's proficiency in each language. However, bilinguals have been shown to both produce (Caramazza et al., 1973; Macleod & Stoel-Gammon, 2009; Simonet, 2016; Sundara et al., 2006) and perceive (Antoniou et al., 2012; Garcia-Sierra et al., 2009) sounds according to the language context of the present surrounding environment.

Work from perception also shows that bilinguals' speech perception processes can be conceptually cued by language identity: they categorize phonemes differently when they are told that they are hearing English as opposed to Spanish or French (Gonzales et al., 2019). This result suggests both that phonetic information is integrated into conceptual language-specific categories, and that expectations can preferentially activate biases in perception. This relationship between language expectation, acoustic form, and representation also interacts in the opposite direction, as some evidence suggests that bilingual speakers can use the acoustic detail of speech to cue language-specific comprehension. Ju and Luce (2004) found that when Spanish-English bilinguals heard Spanish words spoken with an English-like initial phoneme (i.e. a stop consonant produced with voice onset time using English norms rather than Spanish norms), they considered English words sharing the same English onset. However, these English distractor words were not considered any more than distractor items with no phonological onset similarity in either language when those Spanish target words were produced in a Spanish-like way. Thus, fine-grained acoustic-phonetic information can differentially activate a bilingual's two languages.

Phonetics of code-switching

Although it is primarily studied at the syntactic and lexical levels, code-switching has implications for the phonetic features of language. As has previously been discussed, simply being in a dual-language mode (rather than a single-language mode) may influence how bilingual speakers produce words and sounds (i.e. Antoniou et al., 2010; Flege, 1991; Olson, 2020), as may moving between two countries which use different languages (i.e. English in the United States and Portuguese in Brazil; Sancier & Fowler, 1997). However, potentially more nuanced patterns of speech production can be found if we focus on the utterances which contain a code-switch. By considering the case of code-switched utterances specifically, rather than simply interactions that use or could use multiple languages, we are able to focus our analysis on how multiple representation categories interact in production and perception in the moment with minimal adjustment time. This work has used a variety of methods, including cued language-switching, sentence reading, and analysis of naturally-occurring speech. Findings suggest that speech production is different very close to the point of the code-switch than at further-removed locations even within a dual-language context. Production of individual phonemes may shift sub-phonemically, with speech sounds produced in a way that differs phonetically from their unilingual-context norms but not phonologically (Bullock & Toribio, 2009). For example, English voice onset times may be several milliseconds shorter near an English-Spanish code-switch, but not so short that they are better captured by the distribution of VOTs in Spanish tokens by the same speaker. However, the exact roles of language dominance, language identity, acoustic feature, task, and type of code-switch or language-switch are varied and remain somewhat unclear.

In a cued language-switching paradigm, individuals are shown a picture and told to name that picture. The language they should use in naming the picture is cued using a trained symbol such as a colored box or a country's flag. This paradigm allows researchers to compare individuals' responses on "same" trials, in which the same language is used in two consecutive pictures, and "switch" trials, in which a different language is used than in the preceding trial. Goldrick, et al. (2014) found that for Spanish-English bilinguals, the production of English stop consonants is different in switch trials than in stay trials: English voiced and voiceless stops are more Spanish-like in switch trials, especially when producing cognates. Similarly, Olson (2013) tested both English-dominant and Spanish-dominant bilinguals and found evidence that sound contrasts between the two languages were smaller in switch trials than in stay trials. However, in that case the effect of trial type seemed to be driven by accommodation in the dominant language: English-dominant bilinguals produced English that was more Spanish-like in the switch trials, while Spanish-dominant bilinguals produced Spanish that was more English-like. Further, Olson (2013) found an additional interaction with language context such that these patterns occurred only in a mostly-monolingual context with a small proportion of switch trials, and not in a balanced context with an equal number of trials in each language. Comparable effects of language dominance and cued switching extend to Cantonese-English bilinguals (R. K.-Y. Tsui et al., 2019), with unbalanced bilinguals producing a weaker contrast in switch trials in their dominant language while balanced bilinguals show no effect of trial type. However, contrasting evidence from Gavino & Goldrick (2023) shows phonetic effects of language mixing (being in a dual-language task) but not switching.

Sentence-reading paradigms allow us to assess possible effects of planning, as participants can see what word or language they will be asked to produce next and may therefore actively plan for upcoming speech, unlike in the cued switching paradigm. They also consider speech within a sentence context, which may contain intonational contours and aspects of coarticulation that are absent when simply producing single words. Results from Greek-English bilinguals living in Australia suggest that stop consonants in English words are produced with shorter VOTs in Greek sentence frames than in English sentence frames; the reverse, however, is not true for stop consonants in Greek words within English sentence frames (Antoniou et al., 2011). Similar asymmetric effects are found in English-Spanish bilinguals, and crucially appear invariant to the language context beyond the code-switched sentence itself (Olson, 2016). Further, Bullock et al. (2006) show that this asymmetric phonetic transfer may be driven by language identity, language dominance and the position of the target phoneme relative to the code-switch. Spanish-dominant bilinguals produced English stop consonants both preceding and at the point of a code-switch with a shorter VOT compared to how they produced English in unilingual sentences, but production of their Spanish was unaffected by code-switching. English-dominant bilinguals showed the same effects in English and also produced Spanish with a slightly shorter VOT preceding a Spanish-to-English code-switch. While this effect in English follows the convergence which is typically found in this literature, the effect in Spanish is actually one of exaggeration, where the Spanish VOTs of English-dominant bilinguals were more Spanish-extreme immediately preceding a code-switch into English.

A more ecologically-valid way to examine phonetic transfer in code-switching is to analyze naturally-occurring speech, which adds a discourse context and allows participants to control aspects of sentence structure and word choice, including the syntactic position of code-switching. Such designs have provided evidence for this cross-language phonetic transfer, though the extent to which this transfer is asymmetric across languages and whether it primarily occurs in anticipation of a code-switch or due to perseverating activation remains unclear. An analysis of the New Mexico Spanish-English bilingual corpus of speech among Spanish-English bilinguals living in the Southwest of the United States (Cacoullos & Travis, 2018; Travis & Cacoullos, 2013) shows that English voiceless stops were produced with a shorter, and therefore more Spanish-like, VOT when produced temporally closer to Spanish and longer when produced farther from a switch (Balukas & Koops, 2015). The VOTs of Spanish stops are not affected by the amount of time since English was spoken. This asymmetry lends additional support for the proposal in Bullock et al. (2006) that English VOTs show greater convergence than Spanish because English voiceless stops, belonging to the 'long lag VOT' typological category, have more room for potential variance whereas Spanish voiceless stops are relatively constrained.

Analysis of a different corpus, the Bangor Miami corpus of spontaneous code-switching among highly balanced Spanish-English bilinguals (Deuchar et al., 2014) focused on English voiceless stops and therefore does not allow us to assess the symmetry of phonetic transfer between English and Spanish (Fricke et al., 2016). However, the authors also determined whether a given token was *preceded* or *followed* by Spanish within the same utterance, in order to assess the nature of this phonetic transfer from Spanish onto English. If English VOTs are produced in a slightly more Spanish-like way due to anticipation of upcoming Spanish, English tokens will be shorter when they are followed by Spanish. If, however, activation of Spanish phonology perseverates after a lexical and syntactic switch to English, then English tokens will be shorter when they are preceded by Spanish. The authors found evidence for anticipatory effects. When assessing only those tokens that were followed by Spanish, tokens that were closer in time to Spanish were shorter than those that were separated by a longer amount of time. No such effect of distance to Spanish was present for the tokens that were preceded by Spanish (preservatory effects).

Despite the bulk of evidence suggesting asymmetric cross-language phonetic transfer, such patterns are not universally observed. In Piccinini and Arviniti's (2015) semi-naturalistic paradigm in which dyads of early Spanish-English bilinguals discuss a provided topic, the VOTs of both English *and* Spanish were shorter in code-switched utterances than in unilingual utterances. They found some corroborating evidence that English VOTs in code-switched utterances may depend on the distance to the code-switch, though this distance was measured in terms of number of words rather than duration, which may have affected results. There is no evidence that these effects differ for phonemes that occur *before* the code-switch as opposed to *after* the code-switch, although this effect of direction was not explicitly tested.

Though the primary cue analyzed with respect to cross-language phonetic transfer has been voice onset time, several individual studies have addressed other acoustic dimensions. In addition to VOT, Fricke et al. (2016) measured speech rate prior to a code-switch (as defined by the average syllable duration leading up to the code-switch or matched non-switch point). Syllables were significantly slower preceding a code-switch, by an average of approximately 16ms. They also found that code-switched utterances had a greater rate of disfluency than non-switched utterances before the switch point. One possible interpretation of these results is that speakers switch to the other language when they are experiencing a moment of lexical access difficulty. An alternate possibility is that speakers intend to switch languages for a different idiosyncratic reason, and that this internal plan to switch languages is costly to the speech production system, which results in an increase in disfluencies and slightly slower speech rate. Regardless, this analysis of natural speech suggests that suprasegmental aspects of speech show subtle effects of code-switching.

Some aspects of pitch and duration may also be affected by code-switching, including segment lengthening and pitch hyper-articulation of code-switched tokens (Muldner et al., 2019; Olson, 2012, 2016) and a convergence of prosodic contours between the two languages (Muldner et al., 2019; Piccinini & Garellek, 2014). Vowel quality has rarely been studied in code-switching contexts, though some work suggests that some French vowels may be produced differently in French-English than French-only contexts (Muldner et al., 2019).

Patterns of cross-language transfer may also apply to phonological processes in addition to phonetic detail. Olson (2019) exploited two phonological processes whose patterns differ in English and Spanish. English uses perseverative voicing assimilation of the /s/

phoneme, with /s/ produced as [z] following a voiced consonant. Spanish primarily uses anticipatory voicing assimilation with /s/ produced as [z] *preceding* a voiced consonant, though the realization of this pattern is somewhat variable in Spanish. In a sentence reading paradigm including trials with a code-switch appearing directly after a word-final /s/, /s/ in Spanish words was produced with less voicing in switch trials than stay trials, demonstrating a shift toward the English phonological norms (or that a code-switch results in a boundary across which assimilation does not apply). Similar patterns were also found for intervocalic vocalic lenition in L1-Afrikaans/L2-Spanish bilinguals living in Patagonia, Argentina: Spanish voiced stops were produced with less lenition in the presence of Afrikaans than in a unilingual testing mode, while Afrikaans voiced stops did not undergo the Spanish-like lenition even in the presence of Spanish (Henriksen et al., 2021).

Interaction between comprehension and the phonetics of code-switching

Previous work has considered the role of this natural phonetic transfer at the point of a code-switch in comprehension (i.e. Fricke et al., 2016; Piccinini & Garellek, 2014; Shen et al., 2020). This research suggests that for bilingual listeners fluent in the two target languages, hearing natural cues to code-switching aids comprehension.

Fricke, et al., (2016), investigated this question using the visual world paradigm. Spanish-English bilingual listeners' eye movements were recorded while they saw a grid of four images and heard a sentence with the English frame "Click on the picture of the [target]", where the target item was named in either English (creating a unilingual condition) or Spanish (creating a code-switching condition). In the code-switching trials, the images consisted of the target item, an interlingual distractor whose onset matched the target item if the language had stayed the same (i.e. target item PATO (duck) – interlingual distractor POT; target item LIBRO (book) – interlingual distractor LEAF), and two items which were unrelated both semantically and phonologically to the target and interlingual distractor. That is, if participants expected the sentence to continue in English, they would initially look more to the interlingual distractor than the target image because only the interlingual distractor had an onset that matched the target word in English. Half of the code-switching trials were presented without natural cues to code-switching as they had been recorded in an English-only language mode and the target word was cross-spliced into the sentence frame. The other half of the code-switching trials contained naturalistic anticipatory cues to code-switching including a relatively slowed speech rate in the "picture of the" section of the target frame and shorter VOTs for the /k/ in "click" and the /p/ in "picture."

Though participants looked at both the interlingual distractor and the target image before settling on the target in both the cued and uncued conditions, looks to the target relative to the interlingual distractor were higher in the cued condition compared to the uncued condition. This finding suggests that when participants received an anticipatory cue that the language they were hearing was about to switch from English to Spanish, they could predict that switch into Spanish and were thus better able to suppress activation of English representations and focus on items whose names matched the acoustic signal they were hearing *in Spanish*.

In addition to duration-based cues such as VOT and speech rate, listeners may also use intonation contours to anticipate an upcoming code-switch (Piccinini & Garellek, 2014; Shen et al., 2020). Spanish-English bilinguals who heard unilingual and code-switched sentences in both Spanish and English presented with white noise in the background performed above 70% accuracy at transcribing keywords even in a -6 dB SNR. When the target word was in English, these L1-Spanish but English-dominant bilinguals displayed no difference in transcription accuracy between the unilingual and code-switching sentence conditions (Piccinini & Garellek, 2014). All of these sentences were presented as natural recordings, and therefore the code-switching sentences contained natural phonetic cues to code-switching. It is thus impossible to determine whether listeners would have performed just as well on the code-switching sentences if they had been cross-spliced to avoid such cues. However, an acoustic analysis of the stimuli found that the pitch contours of the sentences corresponded to both the language of the initial portion of the sentence and that of the end. That is, at the end of a sentence, F0 contours for English-only and Spanish-only sentences were noticeably different and were appropriate given established intonation norms in each language. Sentence-final F0 contours for the code-switched sentences were intermediate, but with Spanish-to-English sentences being more similar to English-only sentences and English-to-Spanish sentences more similar to Spanish-only sentences. At the beginning of the sentence, the two code-switched sentences still had relatively intermediate F0 contours, but their positions were flipped (Piccinini & Garellek, 2014). Thus, intonation contours of the code-switched sentences changed over time, reflecting a sentence-global cue which would have been available to listeners to help them anticipate an upcoming code-switch, proactively manage relative activation of the language representations, and transcribe the keywords appropriately.

Results from Mandarin-English bilinguals paint a similar picture and further suggest that the absence of anticipatory intonational cues to a code-switch do indeed hinder comprehension of that switch. Using a similar visual-world paradigm as Fricke, et al., (2016), Shen, et al. (2020) presented Mandarin-English bilinguals with English sentence frames which had either an English or Mandarin target word in either sentence-medial or sentence-final position (i.e. “He noticed the [target; English or Mandarin] on the floor”; “He placed the wallet near the [target; English or Mandarin]”). Half of the code-switching trials were cross-spliced to avoid the presence of anticipatory phonetic cues to code-switching; a Mandarin target word from an originally code-switched sentence was spliced into an originally unilingual English sentence frame. Similarly in the English-only sentences, a spliced condition was created by identity-splicing English target words into originally English-only sentences. Looks to the target object as compared to a same-language phonological distractor and an interlingual phonological distractor were analyzed in each condition (unilingual vs code-switched; spliced vs unspliced; sentence-medial vs sentence-final). For the sentence-medial code-switched sentences, listeners were more accurate in looking to the target object in the unspliced (natural) condition than in the spliced (no anticipatory cues to code-switching) condition. Listeners also looked less to the Mandarin competitor object in these same sentence-medial spliced code-switched sentences than in the sentence-medial natural code-switched sentences. This difference in looks to the competitor suggests that presence of anticipatory cues to code-switching affected prediction of whether a code-switch would occur; listeners seemed to expect the language to continue in English without such anticipatory cues (i.e. in the spliced condition) and therefore did not consider items whose onsets matched the onset of the target word *in Mandarin.* However, this effect was not true when the target word was sentence-final rather than sentence-medial. The reasoning for this effect of sentence position is somewhat unclear, though one possibility (explained below) is that the availability of requisite information for engaging in meta-reasoning about the task influenced results of the sentence-final trials.

The findings in these three studies were characterized by small effect sizes, though the results were significant with only 30, 8, and 41 participants respectively. These small effect sizes should not be surprising, as the phonetic cues in the signal that could have allowed them to occur were themselves small. Fricke, et al. (2016) ensured that their cued code-switched stimuli included a 5-10% decrease in VOT and a 2% increase in speech rate relative to the uncued stimuli. Piccinini & Garellek (2014) and Shen, et al. (2020) did not artificially adjust their stimuli in the natural condition but did perform an acoustic analysis comparing the unilingual and naturally code-switched sentence frames. Whereas the sentence-level pitch contour patterns in Piccinini & Garellek (2014) clearly show an anticipatory effect of code-switching, it is unclear whether pitch assimilation (or dissimilation as has been observed in unilingual Mandarin speech; Xu, 1997) is consistently occurring for all Mandarin tones in the English-Mandarin stimuli (Shen et al., 2020). The exact properties of the signal available to participants to modulate expectations of the upcoming language thus remain unclear. That being said, we do not expect a perfect relationship between the magnitude of a change in the stimuli and the magnitude of a resulting effect size in perception, as the signal needs to be passed and transformed through several layers of processing to show up in an experimental design testing speech perception. For example, the difference in the signal needs to be perceptible, this perception of a change needs to be connected to separable representations which could be differentially activated given that change, and that mental activation needs to drive a physical response.

Further, Fricke, et al. (2016) and Shen, et al. (2020) found unexpected effects driven by participants exploiting patterns in experimental design. Spanish-English bilinguals seemed to figure out the relationship between cued code-switching, the target item, and the interlingual distractor. This pattern meant that they were looking to the target more in the cued condition than the uncued condition in the 200ms immediately following target word onset, before they could have programmed and completed an eye saccade in response to comprehension of that word. The predictable nature of the experiment added to this limitation: the cued condition only occurred in code-switched sentences, whereas sentences which continued in English never contained misleading cues as if they would switch into Spanish. Further, all switches occurred at the exact same point in the sentence (after "picture of the"). Therefore, if participants could perceive the changes in speech rate and VOT associated with the cued condition, they could theoretically anticipate an upcoming code switch with 100% accuracy and use the relationship between the labels of the pictured objects in each language to predict the target object. While the sentence frame was slightly less predictable for the Mandarin-English bilinguals as the target item could appear sentence-medially or sentence-finally, they knew that it would eventually appear. When no pictured item occurred prior to the prepositional phrase the target word was therefore necessarily sentence-final; this stimulus pattern contributed to higher accuracy for sentence-final target words in a different experiment with the same participants involving monitoring for a pictured target concept, and could have contributed to this visual world experiment as well.

This meta-awareness of the experimental design can be problematic for our interpretations of the underlying mechanisms driving behavior, as it is difficult to tell what exact aspect of the stimulus drove participants' response. When listeners are given a limited set of possible sentence interpretations and design restrictions allow them to deduce the moment when their comprehension will be assessed, the generalizability of our interpretations about *comprehension* in real-life scenarios is somewhat limited. However, these results do suggest that participants are perceptually sensitive to sub-phonemic cues and that this information has the potential to influence comprehension. As has previously been shown, phonetic detail influences language activation and speech comprehension *and* language-conceptual expectation influences speech perception (Gonzales et al., 2019; Gonzales & Lotto, 2013; Ju & Luce, 2004); this interaction makes it difficult to separate comprehension and expectation from speech perception in previous studies.

Another key component of the discussed studies was the choice of participants; all participants were young adults who had been exposed to both tested languages from an early age and continued to use both languages regularly, including code-switching when talking to bilingual friends and family. By testing frequent code-switchers' perception of the cues to code-switching which are present in natural speech among their general community, the authors were able to assess whether those cues could serve a functional purpose within that community. However, doing so makes it impossible to separate the effects of experience producing these particular phonetic cues, knowledge of the two target languages allowing for the tracking of detailed distributions of particular phonemes conditional to the language being spoken, and previous experience hearing these same cues which could build expectations for how to utilize similar input in an artificial lab-based environment. As described in the Production-Distribution-Comprehension model (MacDonald, 2013), these processes may essentially feed one another in a loop, making it difficult to disentangle their individual contributions. The present study attempts this disentanglement by separating lived experience hearing and producing similar *types* of acoustic cues from the *exact* acoustic cues used in the study.

Many of these concerns raised above are irrelevant for questions focused on the comprehension of code-switching by bilingual speakers, as such questions center on understanding the processes associated with listening to the two languages an individual knows. However, investigating the relationship between previous experience, cognitive and linguistic development, and general language-perceptual mechanisms is supremely important to understand *learning*. Bilingual language learning can happen at any age; though infants are more likely than adults to be in a situation where they are hearing *two* unfamiliar languages, such an experience is not impossible for an adult (e.g. an English-speaking adult may move to an area of Europe where French, German, and English are all spoken regularly and encounter neighbors or co-workers who hold conversations in French and German without using English).

Although learning is not the focus of the present studies, which concern speech perception of unfamiliar languages, it does motivate this dissertation as a whole. If we want to build a theoretical model of bilingual language development that reflects the true variability in the speech available to learners and the states they may experience later in life as adults who are fluent in two languages, we need to know what acoustic properties listeners use to categorize languages while listening, their sensitivity to acoustic cues to code-switching, and how these skills change both with development and experience.

Selection of languages and speaker

Languages

French and Mandarin Chinese were selected as the target language pair for this set of studies because together they meet several key criteria. From a practical perspective, 1) a set of stories intended for children has been translated into both languages, 2) it was possible to find a person who speaks both languages fluently and was able to record stimuli, and 3) it is possible to find potential participants in the United States who have very limited familiarity with both languages. Furthermore, the languages are genetically distinct and thus have few, if any, cognates, and their sound systems utilize distinct phoneme inventories and other phonological features such as lexical tone. Rhythmically, French is classically described as a "syllable-timed" language. Mandarin's rhythm class is debated and it does not clearly match the durational properties of other "syllable-timed" or "stress-timed" languages, though might be more syllable-timed-like than stress-timed-like, at least in some aspects (Mok, 2009; C. Wang et al., 2023). Crucially, a corpus study suggests that Mandarin differs from French in VarcoC (English and Mandarin have similar VarcoC values), nPVI_V, and %V (Mok, 2009). As both adults' and infants' abilities to discriminate between languages seem to be driven by differences in durational properties, rather than categorical rhythm class, it is likely that such differences should be sufficient for discrimination (though discrimination might be easier for more rhythmically-distinct language pairs) (Gasparini et al., 2021; White et al., 2012, 2016).

Perhaps most importantly, this language pair exemplifies the main phonetic features which have been proposed as cues to code-switching in other language pairs: VOT and intonation contour. French, like Spanish, primarily uses relatively short VOTs without aspiration for voiceless stops. Mandarin stop consonants are phonologically slightly different from English stops, as Mandarin only uses voiceless stops but has an aspiration contrast. Nevertheless, Mandarin aspirated voiceless stops are phonetically very similar to English voiceless stops (which are usually aspirated), and Mandarin may in fact use *longer* VOTs than English (Chen et al., 2007).

Therefore, it is likely that the patterns which have been seen in VOTs during Spanish-English and Greek-English code-switching would also apply to French-Mandarin, assuming that these effects are driven by naturally occurring demands on the speech production system due to the need to manage activation of language representations, rather than a cultural- or community-specific byproduct or identity marker. Further, the acoustic analysis presented in Shen, et al. (2020) of English-Mandarin code-switching suggests that at least for some Mandarin tones, there may be anticipatory effects of the specific Mandarin tone immediately following an English-to-Mandarin code-switch on the preceding English contour. Therefore, we believe that French-Mandarin constitutes a language pair that has the possibility to include phonetic cues to code-switching in multiple dimensions, and which would allow us to feasibly test listeners who could perform the task without referring to semantic knowledge.

In what follows we will begin by describing the French-Mandarin speaker and the creation of naturalistic and unnatural code-switched stimuli, followed by the specific stimuli and experiment design for experiments 1-3.

Speaker

A 36-year-old male speaker of Chinese (both Mandarin and Cantonese), Parisian French, and English (YM)[5] was identified. YM's first language was Cantonese. He started learning Mandarin at age 7 and used it regularly in school and/or work starting at age 10. He started learning French in school at age 4. In YM's experience, native speakers of French always thought that he spoke the language natively and never had difficulty understanding him because of his accent, he never made mistakes in French, and he had native or native-like

[5] Note that YM are not the speaker's true initials, these have been chosen randomly to protect his identity

fluency (C2 according to the Common European Framework of Reference levels). In his experience with Mandarin, native listeners often thought he spoke the language natively and occasionally had trouble understanding him because of his accent, he occasionally made mistakes but usually pronounced words like a native speaker, and he had full professional proficiency in the language (C1 according to the CEFR levels). YM recorded two passages (one in Mandarin, one in French), which were then used to rate his accent in each language by native speakers[6].

Ten native Mandarin speakers (1 male, 9 female) recruited through Prolific, the University of Maryland SONA pool, and word of mouth rated YM's accent in Mandarin. The raters ranged in age from 21 to 30, with a mean of 25.6. 9 of the speakers indicated that their first language was Mandarin and that the primary language they spoke was Mandarin. One participant spoke Mandarin as an L2 and English as an L1, but their dominant language was Mandarin rather than English. All speakers also listed fluency in English, and completed the rating task in English. On average, the raters' current exposure to Mandarin was 81.5% of the time, with a range of 18% - 100%. 7 of the raters said that they had no non-native accent in Mandarin, while 2 of the raters said they had a "very light" non-native accent in Mandarin (one did not answer this question). After listening to YM read the passage in Mandarin, the raters answered 3 questions, indicating their confidence that YM is a native speaker of Mandarin, how often they think YM pronounces words like a native speaker of Mandarin, and guessing YM's native language. The responses to the confidence rating question ranged from *not at all confident* to *very confident*, with a median of *"somewhat confident* that YM is a native speaker of Mandarin" (n=2). The responses to the pronunciation frequency question ranged from

[6] As discussed in Baese-Berk et al. (2020), relying on judgments of native speakers of a language is inherently flawed, as individual listeners may differ in their ability to comprehend accented speech and their responses may be biased by socially-driven judgments about non-native speakers. We primarily used this method as a way to assess the potential variability in how listeners may react to YM's production of each language.

occasionally to *mostly* with a median of "YM *mostly* pronounces words like a native speaker of Mandarin" (n=6). The modal native language response was Japanese (n=3). Other responses included an unspecified variety of Chinese (n=1), English (n=1), and several other East or Southeast Asian languages or Chinese varieties including Korean, Cantonese, and Minnan. Four raters specifically mentioned that YM's tone did not fit Mandarin standards: "Japanese or Korean, because the tone of the words in the end of the sentence is rising, this is usually pronounced by Japanese or Korean speakers"; "Japanese or Cantonese because the flow of the whole speech is very similar as Japanese, especially the last 2 words for each sentences." However, raters did note that much of YM's pronunciation was correct including most segmental sounds. Crucially, no rater thought that YM spoke French or any other non-English Indo-European language, including two raters who listed some knowledge of French but less than 5% current exposure to the language.

Ten native French speakers (8 male, 1 female, 1 non-binary) recruited through Prolific rated YM's accent in French. The raters ranged in age from 21 to 40, with a mean of 28.9. All of the speakers indicated that their first language was French and that the primary language they spoke was French. All speakers also listed fluency in English, and completed the rating task in English. On average, the raters' current exposure to French was 59.3% of the time, with a range of 25% - 90%. 9 of the raters said that they had no non-native accent in French, while 1 of the raters said they had almost no non-native accent in French. After listening to YM read the passage in French, the raters answered the same 3 questions. The responses to the confidence rating question ranged from *somewhat confident* (n=1) to *very confident* (n=4), with a median of "*mostly confident* that YM is a native speaker of French" (n=3). The responses to the pronunciation frequency question ranged from *occasionally* (n=2) to *always* (n=4) with a median of "YM *mostly* pronounces words like a native speaker of French" (n=4). The modal native language response was French (n=5). Other responses included English (n=1), French with some sort of caveat (n=2), and an unspecified East Asian language (n=1). The speakers listed as "French with a caveat" thought that YM sounded highly fluent in spoken French but that his accent differed from theirs slightly in ways that they could not quite identify or articulate, such that he could be from Switzerland or Belgium. The one speaker who identified that YM speaks an East Asian language said "Chinese or Japanese or another Asian country. I can't put a clear explanation why, maybe the intonation; I can hear that the person is not a native speaker but I find it difficult to clearly explain why."

Based on these responses it was determined that YM fit the requirements for participation. Although YM's tone in Mandarin did not always fit typical Mandarin standards and may have been more similar to Cantonese (his native language), YM's production in Mandarin was otherwise deemed native-like by the external raters. YM's production of French, similarly, was judged to not quite fit Parisian French standards but raters still mostly believed that French was his native language and picked up on very few times when he did not produce words like a native speaker. The primary goal in selecting a speaker was to find somebody who could speak both Mandarin and French with a high degree of proficiency and did not speak obviously French-accented Mandarin or Mandarin-accented French. YM meets these criteria.

Description of the stimuli used

Six passages were selected from Global Storybooks (https://globalstorybooks.net/) for YM to record. Global Storybooks provides children's stories sourced from countries and cultural traditions around the world translated into over 100 languages. Passages ranged from 19-36 sentences, with a mean of 25.2 sentences. 5-8 switch points were identified in each passage. All switch points were between sentences to ensure that they were a valid location for a code-switch in both languages (unlike for Spanish-English code-switches, detailed corpus information regarding the common syntactic positions for intra-sentential French-Mandarin code-switches is not available). Further, switch points were selected such that there would be a minimum of five seconds between subsequent switches. Four versions of each passage were constructed: entirely in French, entirely in Mandarin, alternating languages at each switch point starting with French, and alternating languages at each switch point starting with Mandarin. YM recorded the two unilingual versions of each passage once each, and the two alternating versions of each passage twice each, for a total of six recordings of each passage. The language-switch points were not marked in the unilingual scripts provided to YM. YM was instructed to read each passage naturally and fluently in a quiet space without distractions and to avoid pausing at the language-switch points in the code-switching stories as much as possible.

Adobe Audition was used to edit all recordings. Each passage was filtered using the built-in de-hum and de-noise filters to remove background noise. Pauses were removed and replaced with a period of silence randomly varying between 100 and 500 milliseconds. All switch points were identified and marked in both the unilingual and switching recordings, and a period of silence was added at each switch point so that all splicing would occur at a zero-amplitude point. All recordings were set to the same average RMS amplitude and amplitude variation within a recording was minimized especially at switch points.

Two stimulus conditions were constructed using these recordings, to create one condition that included acoustic cues to code-switching (the naturalistic condition) and one which did not (the unnatural condition). In the unnatural condition, clips from the unilingual-French and unilingual-Mandarin recordings were interleaved to create passages switching between the two languages but with no natural acoustic cues indicating a switch. In the naturalistic condition, French clips from one of the language-switching recordings were interleaved with Mandarin clips from the other language-switching recording of that same language order (i.e. French-first or Mandarin-first). In both conditions, a version starting with French and a version starting with Mandarin were created. All language switches took place at the designated switch points. Splicing together clips from two language-switching recordings to create the naturalistic condition allowed us to use splicing in both conditions. Therefore, any difference between the two conditions should be due to the language context during the recordings, rather than an unintended editing by-product. In this manner, a total of 24 stimulus passages were constructed with four passage conditions per story: naturalistic French-first, naturalistic Mandarin-first, unnatural French-first, and unnatural Mandarin-first.

All passages were again set to the same average RMS amplitude and if necessary, amplitude was adjusted at switch points such that the transition sounded natural rather than spliced. Every switch point was set to contain 300 ms of silence; other silence periods between sentences varied from 100-500 ms such that the silence at a switch point was not outlyingly short or long and therefore could not serve as an obvious indication of a switch.

Chapter 3: Perception approach: adults (Experiments 1-2)

Experiment 1

Participants

The final sample consisted of 40 self-described English monolinguals and 40 self-described Spanish-English bilinguals. 10 individuals participated in the study but did not meet the demographic criteria for one of the two language groups as described below. An additional 6 monolinguals and 7 bilinguals completed the study but were dropped from further analysis due to not meeting accuracy benchmarks. Participant age, demographic information, and language background for each group is described in table 1. No participant reported known hearing, cognitive, language, or attention impairments.

All participants rated their familiarity with English, Spanish, French, and Mandarin on a scale of 0-4 (0=highly unfamiliar, 2 = neither familiar nor unfamiliar, 4 = highly familiar) and described how often people switch languages either in the middle of a sentence or between sentences when speaking to them (1 = never, 7 = in every conversation). Participants also listed the languages they speak, rated their fluency in each of those languages on a scale of 1-7 (1 = beginning learner, 7 = fluent, native speaker), and estimated the proportion of time they spend speaking each language and listening to each language.

Any participant who reported familiarity in either French or Mandarin of a 3 or 4 was dropped and their data was not analyzed, as those participants may have been using a different strategy to complete the task. To be considered in the monolingual group, participants must have reported a fluency less than 4/7 in all languages other than English and a familiarity with Spanish less than 4/4. To be considered in the bilingual group, participants must have reported a fluency of at least 4/7 in both English and Spanish and less than 4/7 in all other languages, and a familiarity with Spanish of either 3/4 or 4/4.

Monolinguals and bilinguals did not statistically differ in age (t(75.95) = -0.128, p = 0.90), English fluency (t(39) = -1, p = 0.32), English familiarity (all participants reported English familiarity of 4/4), French familiarity (t(74.75) = 1.09, p = 0.28) or Mandarin familiarity (t(67.95) = 0.866, p = 0.39). They did differ in Spanish fluency (t(63.64) = 33.36, p < 0.001), English % exposure (t(41.32) = -11.89, p < 0.001), Spanish % exposure (t(41.47) = 11.96, p < 0.001), code-switching frequency (t(56.68) = 8.20, p < 0.001), and Spanish familiarity (t(51.35) = 14.16, p < 0.001), such that bilinguals were more fluent in Spanish, had less English exposure and more Spanish exposure than monolinguals, and heard code-switching more frequently than monolinguals. Interestingly, 7 English monolinguals listed hearing code-switching at least some of the time. Though these participants had low proficiency in all languages other than English, some of them did speak another language (including Spanish, but not including French or Mandarin) at a beginner or early intermediate level and may have heard code-switching between those languages a small proportion of the time. The maximum code-switching frequency reported by a monolingual was 4/7, by one person.

Table 1: Demographic characteristics of monolingual and bilingual participants in experiment 1

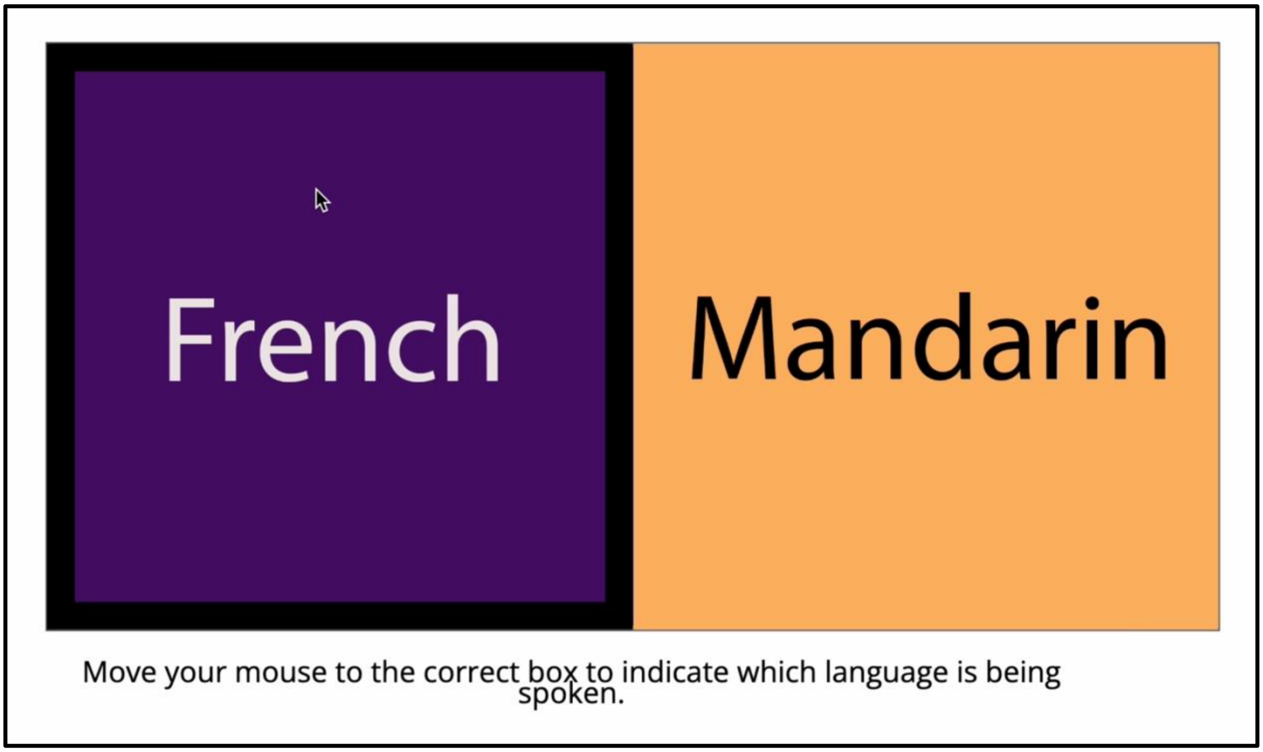
	Monolinguals	**Bilinguals**
Age	31.90 (sd = 7.21)	31.68 (sd = 8.50)
Gender	19 female, 21 male	21 female, 18 male, 1 non-binary
English fluency (max = 7)	7 (sd = 0)	6.98 (sd = 0.16)
Spanish fluency (max = 7)	0.13 (sd = 0.56)	5.93 (sd = 0.94)
English % exposure (speaking and listening)	99.34% (sd = 3.24%)	73.45% (sd = 13.05%)
Spanish % exposure (speaking and listening)	0.63% (sd = 3.24%)	26.29% (sd = 12.84%)
Code-switching frequency (max = 7)	1.30 (sd = .69)	3.33 (sd = 1.40)
English familiarity (max = 4)	4 (sd = 0)	4 (sd = 0)
Spanish familiarity (max = 4)	1.38 (sd = 1.00)	3.8 (sd = 0.41)
Mandarin familiarity (max = 4)	0.2 (sd = .41)	0.3 (sd = 0.61)
French familiarity (max = 4)	.53 (sd = .64)	0.7 (sd = 0.79)
Self-rated understanding of instructions (max = 6)	6 (sd = 0)	6 (sd = 0)
Self-rated effort while completing task (max = 6)	5.93 (sd = .27)	5.975 (sd = 0.16)

Self-rated level of attending to task (max = 6)	5.95 (sd = .64)	5.875 (sd = 0.33)
Self-rated difficulty of task (max = 6)	3.48 (sd = 1.2)	3.6 (sd = 1.30)
Number of trials that passed all accuracy benchmarks (max = 12)	11.23 (sd = 1.10)	10.95 (sd = 1.55)

Procedure

Participants were recruited from Prolific using separate participant filters for each group. Potential participants were then directed to Cognition.run to complete the study, which was programmed using jsPsych (de Leeuw, 2015; de Leeuw et al., 2023). Participants provided informed consent on the Cognition platform. A custom jsPsych plug-in was created which was designed to play an audio file, display an image, and record cursor position relative in time to when the audio file started playing. The image displayed was a box (twice as wide as it was tall), split vertically such that the left and right halves were squares of equal size. One half was labeled Mandarin and the other half was labeled French. Each participant saw the same image throughout all trials of the study; the side-language mapping was counterbalanced across participants. When the participant's cursor was in the left box, a black border appeared around that box as a visual signal to the participant, the same was true for the right box. Figure 1 shows an example of what participants saw during the task when their cursor was in the left box. Note that the two boxes did not fill the entire screen, such that a participant's cursor could have been outside of both boxes.

Figure 1: Example of participant view in experiments 1 & 2



The experiment started with a window resize frame so that the image appeared in the same absolute dimensions for all participants regardless of the size of their device's screen. Participants then listened to YM tell the same short story in both French and in Mandarin and were told the identity of the language while hearing it, in order to familiarize them with both YM's voice and some general properties of each language. Previous research testing language discrimination with adults showed that their performance improved following exposure to one of the languages in question (Bond, et al., 1998). As the goal of this task was specifically to test the effect of natural acoustic cues to code-switching on adults' ability to detect a switch between two languages, rather than their ability to name the language they are hearing with no outside influence, this component of the task should not have affected the outcome of the results, though it may have increased overall accuracy.

Participants also completed a practice trial, which was not scored and for which participants did not receive feedback. The only purpose of the practice trial was to allow participants to become comfortable with the mechanics of the study. In this trial participants listened to an audio file that included clips of a woman telling a story in English interleaved with music, and they moved their mouse to indicate which type of stimuli was playing. Following this language familiarization and practice trial, the full study began.

Two experimental lists were created and participants were randomly assigned to one of the two lists such that an equal number of participants completed each list. Each list consisted of 12 test passages, comprising six unique stories each presented twice, once in each condition such that a participant completed six passages in the unnatural condition and six in the naturalistic condition. Participants were never told the identity of the story they were hearing, though some character names (i.e. Khalai, Nozibele) were repeated throughout a given story and therefore could have become familiar to participants over time. Each story was presented one time in the French-initial order and one time in the Mandarin-initial order, such that each participant heard every segment from every story once in French and once in Mandarin. The mapping between initial language and experimental condition was crossed between lists so that across all participants, every story was heard equally in all four versions. The exact stimuli heard in each experimental list are outlined in Appendix I.

Analysis

Performance was scored on the basis of participants' reaction time to identify the correct language after a language-switch and accuracy in correctly identifying a language while it was playing. These values were calculated individually for each segment in a given passage and for each language in a given passage.

Reaction time: The reaction time measurement indicates how long (in seconds) it took for the participant to move their cursor into the correct box following a language switch.

Segments for which the participant's cursor was already in the target box before that language started playing or for which the participant's cursor never entered the target box were excluded from the reaction time analysis. Reaction time was calculated for each segment and was then averaged across all segments of each language for a given passage to calculate each participant's average French RT and average Mandarin RT for each passage.

Accuracy: The accuracy measurement indicates the proportion of time the participant's cursor was in the target box while each language was playing. To calculate this value for a given segment, we divided the amount of time the participant's cursor was in the target box by the length of the segment. Segment accuracy values thus range from 0 (when the participant's cursor never entered the target box) to 1 (when the participant's cursor was in the target box the whole time). Segments with an accuracy of either 0 or 1 are necessarily excluded from the RT analysis, as the cursor never moved into the target box from another location during those segments. Passage accuracy was calculated separately for each language by dividing the amount of time that the cursor was correctly in that language's box by the length of that language during the passage.

Because the goal of this study was to evaluate listeners' ability to use sub-phonemic acoustic cues to code-switching to aid their identification of a language switch, rather than their ability to recognize and categorize two unfamiliar languages, several accuracy checks were implemented and participants and trials with low accuracy were excluded from further analysis. Only trials (passages) for which all three of these accuracy checks were true were retained in the final dataset: overall passage accuracy (combined across both languages) of at least 55%, passage accuracy for each language of at least 30%, and average passage reaction time for each language under six seconds. The 55% overall accuracy benchmark was selected to ensure that participants were selecting the correct passage at above chance accuracy. The 30% passage accuracy for each language benchmark was selected to ensure that participants moved their mouse into both boxes and did not stay in one box for the entire passage, which would allow them to reach 100% accuracy in one language and 0% accuracy in the other while potentially meeting the 55% accuracy benchmark as passages were necessarily not perfectly balanced for the length of each language. The six-second reaction time benchmark was selected to ensure that participants remained focused and on-task throughout a passage such that our measured reaction times truly reflect the amount of time needed to recognize a language switch and motorically respond to it. A participant needed to reach these minimum benchmarks for more than half of the passages (at least 7 out of 12) to ensure that they contributed enough data to the sample. 13 total participants were dropped from further analysis due to their overall accuracy (6 monolingual and 7 bilingual). The remaining 80 participants contributed an average of 11.09 passages; 73 passages from these remaining participants were dropped due to low accuracy and/or high reaction time.

Data were analyzed at the passage level rather than segment level to reduce variability. To test if there is an effect of natural cues to code-switching, I test if there is a difference in reaction time to identify a switch between the unnatural and naturalistic conditions. As participant group (monolingual vs. bilingual) is strongly correlated with code-switching frequency, these two factors were not included in the same model. Instead, one set of models was fit using language group and a different set of models using code-switching frequency. To test if there is an effect of prior experience with code-switching on the effect of recording condition, I test if there is an interaction between condition and either participant group or code-switching frequency. I also test whether the language (Mandarin or French) interacts with any of these effects, and control for the effects of Mandarin familiarity, French familiarity, and age as participant-level variables.

Each of these fixed effects (language, condition, participant group) were contrast coded (-1 and 1). The four participant-level variables (familiarity with French and Mandarin, code-switching frequency, age) were grand-mean centered. Thus, our estimated coefficient represents deviation from the mean for all factors.

The maximal hierarchical linear model would include random intercepts for participant identity and passage identity, random slopes for language (Mandarin or French) and condition (unnatural or naturalistic) nested within participant and passage, as well as a random slope for group (monolingual or bilingual) nested within passage, resulting in a random effects structure with the following formula: (Language*Condition | Participant) + (Group + Language*Condition | Passage). However, this maximal model failed to converge. Random effects were removed starting with the interaction between language and condition until a maximal convergent model was reached for each analysis. Models were fit and p-values were estimated using the *lmer* function from the R packages lme4 and lmerTest (Bates et al., 2015; Kuznetsova et al., 2017), which uses the Satterthwaite approximation. P-values less than 0.05 are interpreted as statistically significant.

Results

Descriptive Summary:

Before discussing the results from the statistical analysis, I introduce a summary of reaction times and overall accuracy for each language and condition. Monolinguals' average reaction time was 2.31 seconds (sd = 1.04) collapsed across conditions and bilinguals' average reaction time was 2.39 seconds (sd = 1.05) collapsed across conditions. Monolinguals identified the correct language in the unnatural condition in an average of 2.31 seconds (sd = 1.04), and in the naturalistic condition in an average of 2.31 seconds (sd = 1.03). Based on this descriptive analysis, it appears unlikely that monolinguals' reaction time differed based on stimulus condition. Bilinguals, however, responded to the unnatural condition in an average of 2.44 seconds (sd = 1.10), and to the naturalistic condition in an average of 2.34 seconds (sd =

1.01). Descriptively, this suggests that the bilinguals were slower in the unnatural condition than in the naturalistic condition. Statistical analysis, described below, assesses this pattern.

Accuracy describes the proportion of time that participants spent correctly identifying a given language, while that language was playing. Monolinguals were accurate an average of 77.36% of the time (sd = 11.37%) while bilinguals were accurate an average of 76.43% of the time (sd = 11.45%). When split by condition, monolinguals correctly identified the language in the unnatural condition 77.44% of the time (sd = 11.32%) and the naturalistic condition 77.27% of the time (sd = 11.44%). Bilinguals correctly identified the language in the unnatural condition 76.25% of the time (sd = 11.36%), and 76.81% of the time in the naturalistic condition (sd = 11.55%). Accuracy results do not appear to show any effect of participant group or condition, though more detailed statistical analysis follows.

I begin by assessing the distribution of reaction times and accuracy, for both the bilingual and monolingual groups, to ensure normality. Visual assessment of the distribution of the reaction times indicates that they are positively skewed and that the assumption of normality of residuals may not be met. Indeed, the skewness of the raw RT for the monolinguals is 0.98 and for the bilinguals is 0.90. I thus log-transformed the reaction times, which results in normal distributions, with skewness of -0.02 for the monolinguals and -0.18 for the bilinguals.

The accuracy scores are negatively skewed (monolinguals: -1.05; bilinguals: -0.97). This is expected, given the ceiling effects inherent in accuracy data. However, the dependent measure is continuous, unlike most accuracy measures, because I calculated the proportion of time a participant spent identifying the correct language while it was playing, rather than whether a particular response was correct or incorrect (which would be binary) or the number of correct responses (which would be a count outcome). Thus, generalized linear mixed-effects models with link functions that assume binary or count variables are inappropriate. I therefore transformed the dependent variable using the logit transformation (Warton & Hui, 2011), and then used linear multi-level effects models to analyze the responses. Using the logit transformation results in a skewness of -0.26 for the monolinguals and -0.13 for the bilinguals, such that the residuals are expected to also be normally distributed in the final models.

Statistical analysis of reaction time:

The maximal convergent model for an analysis of reaction time using **participant group as a fixed effect** included random intercepts of story and participant, random language-by-participant slopes, and random condition-by-story and language-by-story slopes. The specification of this model is thus **log(RT) ~ language * condition * group + Mandarin + French + age + (language + condition | story) + (language | participant).** Only the effect of age reached significance at an alpha-level of 0.05, such that older participants responded more quickly ($t(75.43) = -1.995$, $\beta = -0.007603$, $p < 0.05$). However, the interaction between condition and participant group was marginally significant ($t(1595) = -1.804$, $\beta = -0.01422$, $p = 0.0715$). This interaction was further explored below by fitting a model separately for bilinguals and monolinguals (see below).

A simplified version of this model was fit which only included those fixed effects which had an estimated p-value less than 0.1, or were included in a higher-level interaction, resulting in a model specification of **log(RT) ~ condition + group + condition:group + age + (language + condition | story) + (language | participant)**. This model did not statistically significantly differ from the full model ($\chi^2(6) = 3.4432$, $p = 0.752$). Both the effect of age ($t(77.32) = -1.925$, $\beta = -0.007324$, $p = 0.0579$) and the interaction between condition and participant group ($t(1597) = -1.801$, $\beta = -0.01419$, $p = 0.072$) were marginally significant at an alpha-level of 0.05. I return to these potential effects below.

Figures 2 and 3 examine the effect of stimulus condition for the monolingual and bilingual participant groups. Figure 2 shows a box-and-whisker plot of all participants' average passage reaction times for each story and compares the naturalistic and unnatural conditions.

Figure 3 shows the proportion of passages with an average reaction time less than or equal to a given value, for all 50-ms intervals between 0 and 6 seconds. In both figures we see that the monolinguals' reaction times do not differ between the naturalistic and unnatural conditions while the bilinguals are slightly faster in the naturalistic condition.

Figure 2: Experiment 1: Reaction time (s) to detect a code-switch by participant group and stimulus condition

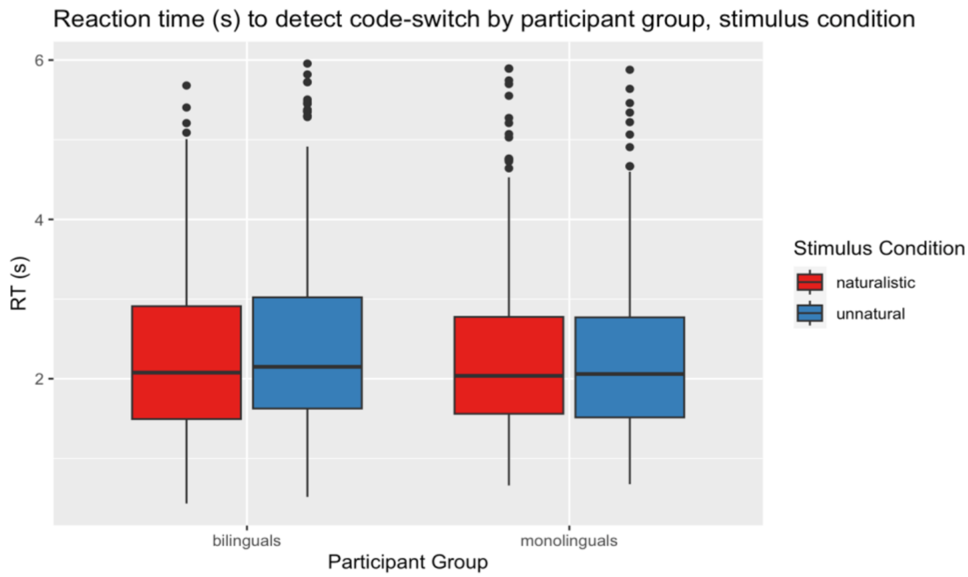


Figure 3: Experiment 1: Time course of identifying the code-switch across all participants and conditions

The maximal convergent model for an analysis of reaction time using **code-switching frequency as a fixed effect** included random intercepts of story and participant, random language-by-participant slopes, and random condition-by-story and language-by-story slopes, resulting in a model specification of **log(RT) ~ language * condition * CSfreq + Mandarin + French + age + (language + condition | story) + (language | participant)**. There was a significant interaction between condition and code-switching frequency ($t(1600) = 2.531$, $\beta = 0.01365$, $p < 0.05$) such that the effect of condition was greater for participants who reported hearing code-switching more often and a significant effect of age ($t(75.47) = -1.993$, $\beta = -0.007585$, $p < 0.05$) such that older participants had generally faster reaction times.

A simplified version of this model was fit using only the fixed effects with an estimated p-value less than 0.1, with a model specification of **log(RT) ~ condition + CSfreq + condition:CSfreq + age + (language + condition | story) + (language | participant)**. This simplified model was not statistically significantly different from the full model ($\chi^2(6)$ = 3.1178, p = 0.7939), indicating that controlling for language or familiarity with Mandarin or French did not improve the fit of the model. The simplified model shows a significant interaction between condition and code-switching frequency (t(1601) = 2.531, β = 0.01364, p < 0.05) and a marginally significant effect of age (t(77.32) = -1.922, β = -007298, p = 0.0582) in the same directions as seen in the full model. Further, this simplified model is a significantly better fit than one that includes the main effects of condition, code-switching frequency, and age, but not the interaction between condition and code-switching frequency ($\chi^2(1)$ = 6.4, p < 0.05), which suggests that including that interaction is important.

Figure 4: Experiment 1: Reaction time (s) to detect a code-switch by participant group, code-switching frequency, stimulus condition

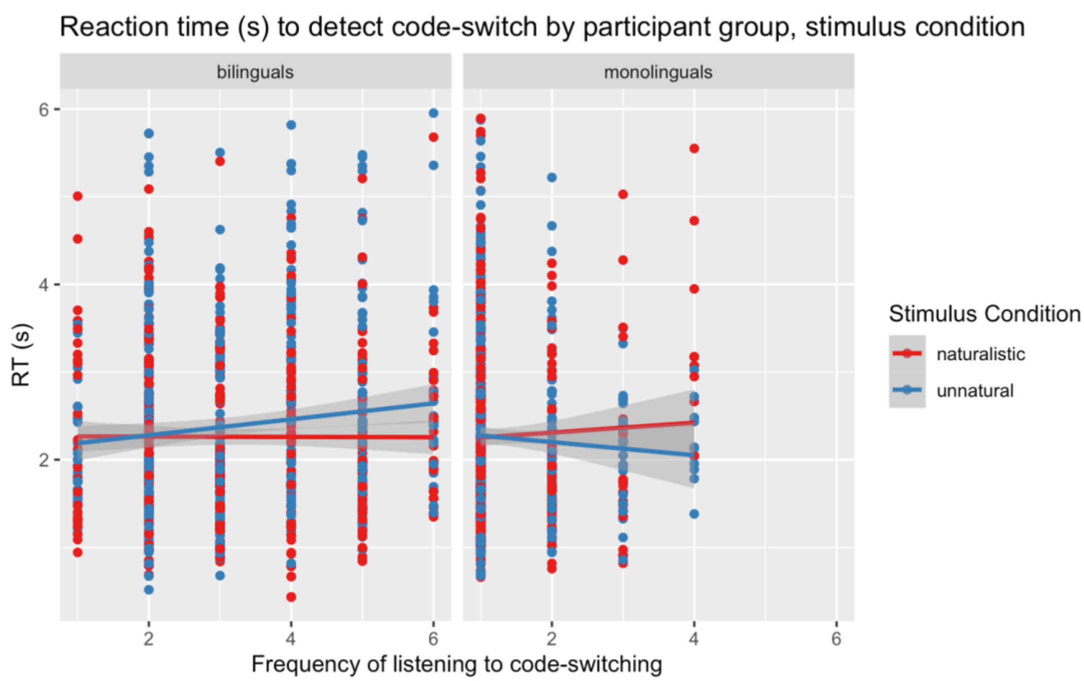
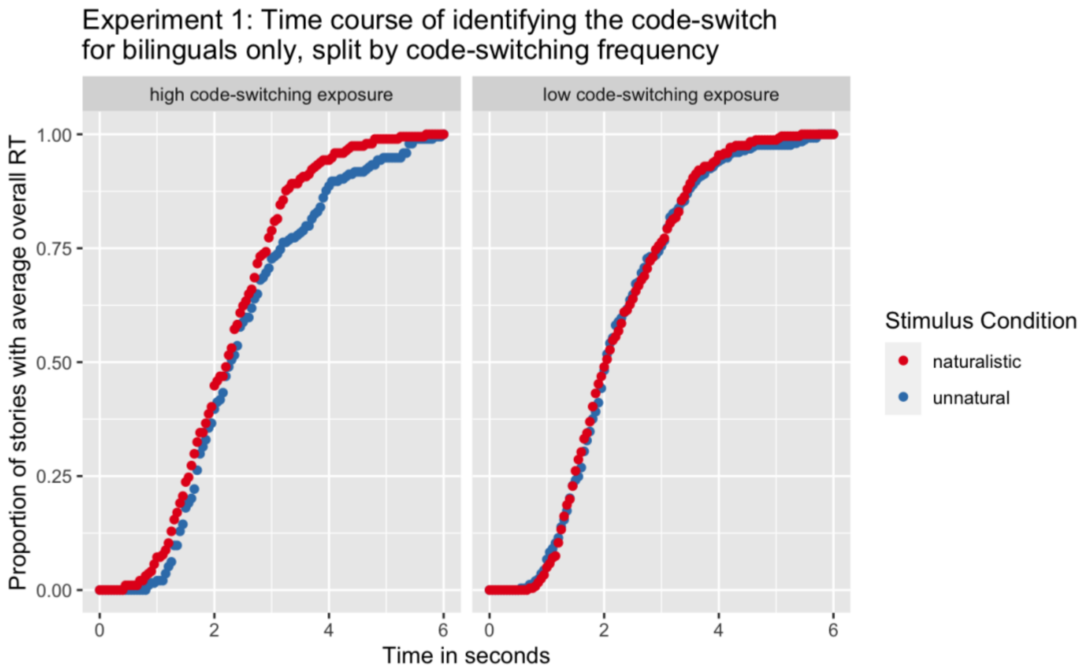


Figure 5: Experiment 1: Time course of identifying the code-switch for bilinguals only, split by code-switching frequency



Given the difference in code-switching frequency between the monolinguals and bilinguals, I fit the code-switching frequency model above separately for the two participant groups in order to better understand the nature of the interaction between code-switching frequency and condition.

The maximal convergent model for the **monolinguals only** contained random intercepts of passage and participant, random slopes of condition and language within passage, and random slopes of language within participant, resulting in a model specification of **log(RT) ~ language * condition * CSfreq + Mandarin + French + age + (Condition + Language | Story) + (Language | Participant)**. The model resulted in no significant or marginally significant fixed effects, suggesting that neither experimentally controlled factors (language, condition) nor individual factors (code-switching frequency, age, familiarity with one or both of the target languages) influenced reaction times for the monolingual participants.

The maximal convergent model for the **bilinguals only** contained random intercepts of passage and participant, random slopes of condition within passage, and random slopes of language within participant, resulting in a model specification of **log(RT) ~ language * condition * CSfreq + Mandarin + French + age + (Condition | Story) + (Language | Participant)**. This is a different effects structure as was used for all previous analyses; a model using random slopes of both condition and language within passage failed to converge. This model resulted in a significant interaction between condition and code-switching frequency ($t(789.786) = 1.982$, $\beta = 0.017$, $p < 0.05$). No other fixed effects approached significance in this model.

The maximal random effects structure that allowed a simplified version of this model containing this significant interaction and the two main fixed effects included in the interaction (condition and code-switching frequency) to converge only had random intercepts of passage and participant, without random slopes (**log(RT) ~ condition * CSfreq + (1 | Story) + (1 | Participant))**. This simplified model also resulted in a significant interaction between condition and code-switching frequency, $t(840.9) = 2.004$, $\beta = 0.17$, $p < 0.05$). Model comparison between these two models showed that they do not significantly differ, $\chi^2(11) = 16.69$, $p = 0.117$, which suggests that the other factors included in the maximal model did not influence the model fit or results. A model excluding that interaction between condition and code-switching frequency was a significantly worse fit than the simplified model including that interaction, $\chi^2(1) = 4.02$, $p < 0.05$.

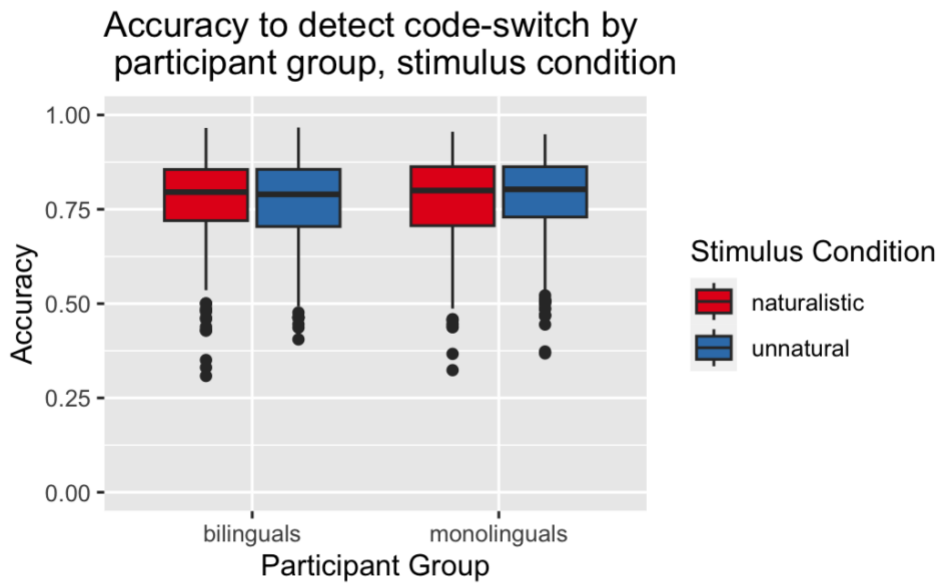
This significant interaction shows that bilinguals who reported more experience listening to code-switching had a greater effect of passage condition than those with less experience listening to code-switching. Specifically, a post-hoc t-test comparing the logs of the reaction times in the two conditions for the bilinguals with high code-switching frequency (greater than the median) was significant, $t(385.5) = 2.5647$, $p < 0.05$, while this was not true for the bilinguals who reported code-switching frequency less than or equal to the median ($t(492) = .064$, $p = 0.95$). The bilinguals who reported hearing code-switching frequently were slower to detect the language switch in the unnatural cues condition as compared to the natural cues condition.

This pattern can also be seen in Figure 4, which show reaction time to detect a code-switch as a function of condition, code-switching frequency, and participant group, and Figure 5, which shows the time course of identifying a code-switch for the bilinguals in each condition, stratified using a median split of code-switching frequency. For the monolinguals, the speed to detect a language switch using naturalistic cues was unaffected by code-switching frequency. However, those bilinguals who heard code-switching more often were slower when they heard unnatural cues as compared to naturalistic ones.

Statistical analysis of accuracy:

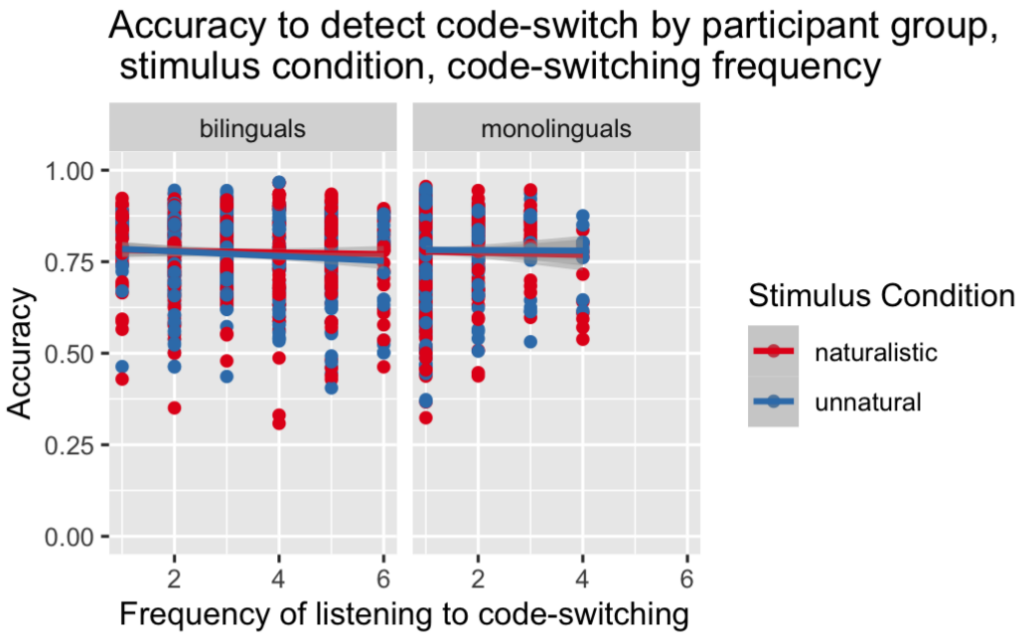
The maximal convergent model using logit-transformed accuracy as the dependent variable and including **participant group as a fixed effect** rather than code-switching frequency included random intercepts of passage and participant, random slopes of condition and language within passage, and random slopes of language within participant for a model specification of **logit(accuracy) ~ language * condition * group + Mandarin_familiarity + French_familiarity + age + (Condition + Language | Passage) + (Language | Participant)**. No fixed effects were statistically significant in this model, though the effect of age was marginally significant ($t(75.06) = 1.883$, $\beta = 0.0094$, $p = 0.0635$), such that older participants may have had slightly higher overall accuracy scores.

Figure 6: Experiment 1: Accuracy by participant group and condition



The maximal convergent model using accuracy as the dependent variable and including **code-switching frequency** as a fixed effect rather than participant group included random intercepts of passage and participant, random slopes of condition and language within passage, and random slopes of language within participant for a model specification of **logit(accuracy) ~ language * condition * CS_freq + Mandarin_familiarity + French_familiarity + age + (Condition + Language | Passage) + (Language | Participant)**. No fixed effects were statistically significant in this model, though marginally significant effects were observed for the interaction between condition and code-switching frequency ($t(1591) = -1.881$, $\beta = -0.01477$, $p = 0.06$) such that the effect of condition was greater for participants who heard code-switching more often and age ($t(75.04) = 1.886$, $\beta = 0.0094$, $p = 0.063$) such that older participants were more accurate. As no effects were significant in this analysis of accuracy, there is no evidence to suggest that condition or previous language experience impacted participants' overall accuracy at identifying which language was being presented.

Figure 7: Experiment 1: Accuracy by participant group, code-switching frequency, and condition



Discussion

The goal of study 1 was to investigate if listeners are sensitive to sub-phonemic acoustic cues to code-switching and can use these cues to aid detection of a language switch. 40 English monolingual listeners and 40 Spanish-English bilingual listeners who reported no more than minimal exposure to French and Mandarin Chinese listened to French-Mandarin code-switching. Participants continually identified the language being spoken by moving their computer cursor, and I evaluated their speed and accuracy at detecting when the language switched. Participants completed this task for six stories that contained natural cues to code-switching, and for six stories which were recorded in a single-language mode and then spliced together such that they did not contain natural cues to code-switching; performance was compared across the two conditions.

I found evidence of an interaction between story condition and previous experience with bilingualism/code-switching in participants' reaction times to identify a code-switch. Spanish-English bilingual adults (particularly those who said that they hear others code-switch frequently) were faster to identify a language switch in the natural than in the unnatural condition, meaning the presence of natural acoustic cues to code-switching aided their performance (or that the absence of such cues hindered their performance). Such an effect of condition was not present for the monolingual listeners. Averaging across conditions, monolinguals and bilinguals showed similar overall accuracy and reaction time at identifying a language switch.

It is unclear based on these results alone whether we should interpret the bilinguals' performance in the natural or the unnatural cues condition as their "baseline". That is, is hearing some acoustic cue that signals a code-switch so common and such a natural cue for the bilinguals that such prediction occurs automatically, but the absence of such a cue leads to an artificial slowdown in detection of a code-switch? Or, alternatively, are bilinguals highly sensitive to natural cues to code-switching and able to use them to detect a switch, but are not necessarily expecting such cues to be present (especially in unfamiliar languages)? One possible way to answer this question may be to compare performance between conditions that vary in the strength of their cues to code-switching. If such cues are helpful, then listeners should be even faster to identify a language change when they are stronger (i.e. greater change in VOT or speech rate, more drastic shift in intonation contour) than when they are present but weak. However, if the absence of cues is harmful then listeners should be slower for unnatural switches than natural switches with strong *or* weak cues, but not necessarily show a difference between the strong and weak conditions. One challenge to this potential follow-up study is that it requires us to have a detailed understanding of the exact phonetic cues that are present in the signal and useful to listeners in order to determine the degree of change in each acoustic dimension that would make a particular instance of code-switching contain strong or weak cues.

One somewhat interesting result from study 1 is that these effects were only present in the analysis of reaction times, but not in the analysis of accuracy. This apparent discrepancy is not particularly surprising, though, due to the time periods over which the two dependent measures take information. The reaction time measure is based on participants' responses immediately after a language switch occurs, and shows us that on average participants know the language presently being spoken within 2.5 seconds. The accuracy measure, however, is based on the entire length of each segment. Segments range considerably in length, to prevent participants from being able to find a pattern in the frequency of code-switching. Some are just over five seconds long, while others are as long as 18 seconds. Previous acoustic analyses of natural code-switching (Balukas & Koops, 2015; Fricke et al., 2016) suggest that the effect of code-switching on voice onset time length interacts with distance, such that tokens produced closer in time to that code-switch point experience more cross-language phonetic transfer than those produced farther away. Therefore, large portions of many segments may be mostly unaffected by the context in which they were recorded, so it is unlikely that condition would affect participants' responses in the middle of those segments. Further, most participants rarely moved their mouse out of a box prematurely. Once they correctly identified a language, they continued doing so for the rest of the segment. The accuracy measure therefore incorporates both the period of time already captured by the reaction time measure, plus a period of 3-15 seconds that is unlikely to be responded to incorrectly. Although the accuracy measure plays an important role in filtering out trials that were particularly difficult for participants and removing them from further analysis, it may not yield enough relevant variability to be useful in addressing my research questions.

Another question that remains about the bilinguals' performance specifically is the exact cues that they were using to speed their identification of the language switch in the naturalistic condition. While previous studies testing bilinguals' use of phonetic cues in comprehension of code-switching conclude that anticipatory cues allow listeners to predict an upcoming switch, these studies essentially limited the available cues to solely those which could be anticipatory. That is, both Fricke, et al. (2016) and Shen, et al. (2020) used sentence frames in a consistent matrix language (English) and single-word language switches into the bilingual participants' other language (Spanish or Mandarin Chinese, respectively). They varied whether the sentence frame preceding that single-word switch could have contained such naturalistic phonetic cues to code-switching. However, that target word, even if it had originally been recorded in a code-switched sentence and could have been produced with perseverative cues to code-switching, likely had less potential informativity than the larger sentence frame as it was only a single word. As previous examinations of the acoustics of code-switching have found evidence for both anticipatory and perseverative effects in natural speech (Balukas & Koops, 2015; Fricke et al., 2016), it is possible that when cues in both directions are available to listeners, both are useful. The present study does not allow us to distinguish between the use of anticipatory and perseverative cues because both are present in the naturalistic condition and absent in the unnaturalistic condition.

Experiment 2

Procedure

Experiment 2 was designed such that a given language switch only has potential anticipatory cues *or* potential perseverative cues. The goal of study 2 was to determine whether bilingual listeners used primarily anticipatory or perseverative cues to quickly identify the language switch in the natural-switches condition in study 1.

The same base recordings as described for study 1 were used for study 2 but rearranged to create new passages. For study 2, in each passage, segments from one language recorded in the natural-switching condition were spliced with segments of the other language recorded in the unilingual condition. By splicing between these two recording types, I created passages in which a given switch between languages had only anticipatory cues (because the first language was from the natural-switching recording and the second language was from the unilingual recording) or perseverative cues (because the first language was from the unilingual recording and the second language was from the natural-switching recording). Every passage therefore contains both conditions, with French in one condition and Mandarin in the other condition. Four new versions of each passage were constructed: 1) Mandarin-first, Mandarin is from the natural-switches recording and French is from the unilingual recording; 2) French-first, Mandarin is from the natural-switches recording and French is from the unilingual recording; 3) Mandarin-first, French is from the natural-switches recording and Mandarin is from the unilingual recording; 4) French-first, French is from the natural-switches recording and Mandarin is from the unilingual recording. In conditions 1 and 2, Mandarin-French switches contain anticipatory cues and French-Mandarin switches contain perseverative cues; the reverse is true in conditions 3 and 4.

As in study 1, all language switches were set to contain a 300ms period of silence between the two languages. Other silent periods between phrases or sentences in a single language vary randomly from 100 to 500 ms. All language segments and passages were set at the same average RMS amplitude. All participants heard a total of twelve passages, six with Mandarin from natural-switches and French from unnatural-switches such that French-to-Mandarin includes perseverative cues and Mandarin-to-French includes anticipatory cues, and six with Mandarin from unnatural-switches and French from natural-switches such that French-to-Mandarin includes anticipatory cues and Mandarin-to-French includes perseverative cues.

If anticipatory changes are present in the speech signal and the primary cue that is useful to listeners, then in study 2 listeners should be faster to identify a given language during code-switches that go from a natural to an unnatural segment, as the initial segment would contain such anticipatory cues. Though listeners must respond with the language of the switched-into segment, which does not contain natural cues to code-switching, the presence of anticipatory cues (and sensitivity to such cues) would allow the previous segment to improve response times.

If perseverative changes are present in the speech signal and the primary cue that is useful to listeners, then we expect the reverse, that listeners will be faster to identify a given language during switches that go from an unnatural to a natural segment. The target segment itself would have perseverative cues to the code-switch at the beginning of the segment due to how it was originally recorded. Thus, listeners would use the initial portion of the current segment to aid identification of the language in that segment.

If instead of the presence of cues providing an added benefit to language identification and speech comprehension, the absence of cues in an unnatural spliced condition actually *hurts* listeners, we would expect the same pattern of results. That is, if the absence of anticipatory cues is particularly harmful, listeners will be faster to identify a language during switches that contain these anticipatory cues than switches that do not (i.e., the ones that instead have perseverative cues). Similarly, if the absence of perseverative cues is particularly harmful, listeners will be faster to identify a language during switches that contain those perseverative cues than those that do not (the switches that instead have anticipatory cues). If anticipatory and perseverative cues work together and both play approximately equal roles in code-switch detection and identification, then there should be no difference in reaction times between the two conditions in study 2.

<u>Participants</u>

A new group of 40 self-reported Spanish-English bilinguals were recruited through Prolific for study 2. Nobody from study 1 also participated in study 2. The procedures for study 2 were the same as in study 1. Participants completed a questionnaire, were familiarized with

YM speaking French and Mandarin, practiced the task, and then listened to 12 passages while moving their cursor to identify whether French or Mandarin was being spoken. Each participant heard all six stories in the Mandarin-first order and all six in the French-first order. All participants heard a given language presented in one condition (unnatural or naturalistic) for six passages and the other condition for the other six passages. As in study 1, which side of the screen was assigned to each language was consistent within a given participant and counterbalanced across participants and the order of all stories was randomized.

Data were scored and both reaction time and accuracy were measured as in study 1. The same trial and participant exclusion guidelines were followed. Twelve individuals who participated were excluded from data analysis due to demographic factors (i.e. high familiarity with French or Mandarin, low Spanish fluency or familiarity). Four additional individuals who met the demographic criteria were excluded due to low accuracy or high reaction time on at least six trials during the study. The remaining 40 participants contributed an average of 11.425 trials each (23 total trials were excluded).

A series of two-tailed t-tests with unequal variance were conducted to compare the demographics of the bilinguals in experiments 1 and 2. A χ^2 test was used to compare the gender distribution across experiments. The groups do not statistically significantly differ in any of these criteria. Table 2 compares demographic details between the two Spanish-English bilingual groups.

Table 2: Demographic characteristics of the bilingual samples in experiment 1 and experiment 2

	Experiment 1 (bilinguals)	***Experiment 2***
Age	31.68 (sd = 8.50)	30.53 (sd = 8.42)
Gender	21 female, 18 male, 1 non-binary	18 female, 19 male, 3 non-binary
English fluency (max = 7)	6.98 (sd = 0.16)	7 (sd = 0)
Spanish fluency (max = 7)	5.93 (sd = 0.94)	6.10 (sd = 0.71)

English % exposure (speaking and listening)	73.45% (sd = 13.05%)	72.39% (sd = 13.42%)
Spanish % exposure (speaking and listening)	26.29% (sd = 12.84%)	27.49% (sd = 13.28%)
Code-switching frequency (max = 7)	3.33 (sd = 1.40)	3.90 (sd = 1.35)
English familiarity (max = 4)	4 (sd = 0)	3.98 (sd = 0.16)
Spanish familiarity (max = 4)	3.8 (sd = 0.41)	3.9 (sd = 0.30)
Mandarin familiarity (max = 4)	0.3 (sd = 0.61)	0.35 (sd = 0.48)
French familiarity (max = 4)	0.7 (sd = 0.79)	0.95 (sd = 0.60)
Self-rated understanding of instructions (max = 6)	6 (sd = 0)	6 (sd = 0)
Self-rated effort while completing task (max = 6)	5.975 (sd = 0.16)	5.95 (sd = 0.23)
Self-rated level of attending to task (max = 6)	5.875 (sd = 0.33)	5.84 (sd = 0.37)
Self-rated difficulty of task (max = 6)	3.6 (sd = 1.30)	3.21 (sd = 1.28)
Number of trials that passed all accuracy benchmarks (max = 12)	10.95 (sd = 1.55)	11.43 (sd = 1.15)

Data were analyzed using the same methods as in experiment 1: I calculated participants' reaction times to identify the correct language and the amount of time they spent identifying the correct language during each segment, for each story. Passages which had an average reaction time greater than 6 seconds for either language, accuracy less than 30% for either language, or total accuracy less than 55% accuracy were excluded. Any participant for whom more than 5 passages was excluded was dropped from further analysis entirely.

<u>Results</u>

Descriptive summary:

I begin with a summary of reaction times and overall accuracy for each language and condition. Participants' mouse correctly entered the box associated with French an average of 2.23 seconds (sd = 0.94) after French began, and entered the Mandarin box an average of 2.25 seconds (sd = 0.97) after Mandarin began. Therefore, it seems unlikely that participants were better at identifying one language than the other language. However, comparing the reaction times of the two conditions within each language suggests that there may be an effect of condition for French but not for Mandarin: participants responded to French in 2.08 seconds when provided with anticipatory cues only, but needed 2.37 seconds when provided with perseverative cues only. In Mandarin the difference was slight: 2.23 vs 2.27 seconds, respectively.

Accuracy tells a slightly different story, though recall the discussion from study 1 that accuracy is likely a less sensitive measure than reaction time in this task, as it incorporates continued responses over a period of 5.5 – 18 seconds, and therefore may not reflect the influence of cross-language phonetic transfer at the point of a code-switch specifically. When French was playing, participants' mouse was located in the French box 77.45% of the time (sd = 11.36%), and when Mandarin was playing, participants' mouse was located within the Mandarin box 79.51% of the time (sd = 9.74%). Participants correctly indicated French 78.74% of the time during segments that had anticipatory cues and 76.16% of the time during segments that had perseverative cues, and Mandarin 80.27% and 78.74% of the time, respectively.

In what follows I discuss a statistical analysis of these results. I used multi-level models to assess the effects of language (French or Mandarin), segment condition (recorded unilingually or with natural code-switching), and participant code-switching frequency while accounting for participants' age, Mandarin familiarity, and French familiarity. As in study 1, data were analyzed at the passage level rather than segment level to reduce variability. To test if there is an effect of the direction of natural cues to code switching, I test if there is a difference in reaction time to identify a switch that has only anticipatory cues as compared to identifying a switch that has only perseverative cues. I additionally test if this effect interacts with participants' code-switching frequency or language (Mandarin vs French), and control for the effects of Mandarin familiarity, French familiarity, and age as overall participant-level variables. The two fixed effects (language, condition) were effects coded (-1 and 1). The four participant-level variables (familiarity with French and Mandarin, code-switching frequency, age) were grand-mean centered for ease of interpretation.

The maximal hierarchical linear model would include random intercepts for participant identity and passage identity, as well as random slopes for language (Mandarin or French) and condition (anticipatory or perseverative) nested within participant and passage, resulting in a random effects structure with the following formula: (Language*Condition | Participant) + (Language*Condition | Passage). However, this maximal model failed to converge. Random effects were removed starting with the interaction between language and condition until a maximal convergent model was reached for each analysis. Models were fit and p-values were estimated using the *lmer* function from the R packages lme4 and lmerTest, which uses the Satterthwaite approximation. P-values less than 0.05 are interpreted as statistically significant.

As in experiment 1, the reaction times in experiment 2 were positively skewed (skewness = 1.01) and the accuracy measure was negatively skewed (skewness = -1.19). I therefore log-transformed the reaction time data, which leads to a normal distribution (skewness = 0.15) and logit-transformed the accuracy data. This transformed accuracy measure is still somewhat negatively skewed (-0.43), but I refrained from additional transformations which could result in overfitting to this particular distribution, and because the analysis approach is relatively robust.

Statistical analysis of reaction time:

The maximal convergent model for the test of reaction time included random intercepts for participant and passage and a random slope of condition nested within passage, with a model specification of **log(RT) ~ language * condition * code-switching frequency + Mandarin familiarity + French familiarity + age + (1 | Participant) + (Condition | Passage)**. This model resulted in a significant effect of condition such that participants were faster to identify segments that contained anticipatory cues only than segments that contained perseverative cues only ($t(11.28) = -2.35$, $\beta = -0.030$, $p < 0.05$). There was also a significant effect of age such that older participants had overall slower reaction times ($t(34.75) = 2.766$, $\beta = 0.012$, $p < 0.01$). Further, there was a significant interaction between condition and language ($t(846.72) = -2.03$, $\beta = -0.021$, $p < 0.05$). The nature of this interaction is explored further below by conducting separate analyses for French and Mandarin.

A simplified version of this model was fit using only the fixed effects and interactions which had a p-value less than 0.1 in the maximal model, and any main effects which contributed to an interaction but were not themselves significant. This model used the same random effects structure as the full model described above. The reduced and full models do not statistically significantly differ from one another, χ^2 (6) = 6.58, $p = 0.36$, which shows that controlling for code-switching frequency, Mandarin familiarity, or French familiarity did not improve the fit of the model. This reduced model resulted in the same pattern of significant effects as the full model: significant main effects of condition ($t(11.28) = -2.353$, $\beta = -0.030$, $p < 0.05$) and age ($t(37.78) = 2.711$, $\beta = 0.012$, $p < 0.05$) and a significant interaction between language and condition ($t(849.72) = -2.029$, $\beta = -0.021$, $p < 0.05$). Further, this model including both the main effect of condition and the interaction between language and condition was a better fit than a model excluding the interaction only ($\chi^2(1) = 4.12$, $p < 0.05$) and a model excluding both the interaction and the main effect of condition ($\chi^2(2) = 9.05$, $p < 0.05$).

In order to interpret the interaction between language and condition observed above, I conducted separate analyses for switches into French and switches into Mandarin. For switches **into French**, the maximal convergent model included random intercepts of participant and passage and a random slope of condition for each passage. This model showed a significant effect of condition ($t(393.0) = -3.528$, $\beta = -0.051$, $p < 0.001$) such that participants responded faster to switches that contained anticipatory cues than switches that contained perseverative cues and a significant effect of age ($t(33.57) = 2.138$, $\beta = 0.012$, $p < 0.05$) such that older participants responded more slowly in general. No other effects were significant. A simplified model using the same random effects structure and including only those effects that had an estimated p-value less than 0.1 did not significantly differ from the maximal model, $\chi^2(4) = 4.50$, $p = 0.343$. This simplified model continued to show a significant effect of condition ($t(9.57) = -3.194$, $\beta = -0.051$, $p < 0.05$) and a marginally significant effect of age ($t(37.53) = 2.006$, $\beta = 0.010$, $p = 0.0521$) in the same directions as previously. This model was a significantly better fit than one excluding the main effect of condition, $\chi^2(1) = 7.78$, $p < 0.01$.

Models using a random effects structure with random intercepts for participant and passage, but no random slopes, were fit for switches **into Mandarin**, rather than switches into French. These results did not show a significant effect of condition for switches into Mandarin, though the effect of age was significant such that older participants had generally longer reaction times ($t(33.60) = 3.317$, $\beta = 0.016$, $p < 0.01$). There was also a marginally significant effect of Mandarin familiarity such that participants who were more familiar with Mandarin responded more quickly ($t(33.88) = -1.777$, $\beta = -0.137$, $p = 0.085$). This effect of age remained when assessed using a simplified model ($t(36.51) = 3.405$, $\beta = 0.015$, $p < 0.01$), though the effect of Mandarin familiarity was no longer marginally significant in this simplified model ($p = 0.11$). The fit of that simplified model did not differ from the full model (χ^2 $(4) = 2.07$, $p = 0.722$).

Figures 8-10 show this pattern of results. In Figure 8, participants' reaction time is plotted as a function of the stimulus condition and language. Figure 9 further plots reaction time as a function of age and shows that older participants were consistently slower than younger participants in this task. In Figure 10, the proportion of participants with a reaction time less than or equal to a given value is plotted, in 50-ms increments from 0 to 6 seconds. This time-course plot is split by language and condition and shows a consistent difference between anticipatory and perseverative conditions for switches into French.

Figure 8: Experiment 2: Reaction time (s) to detect code-switch by language, stimulus condition

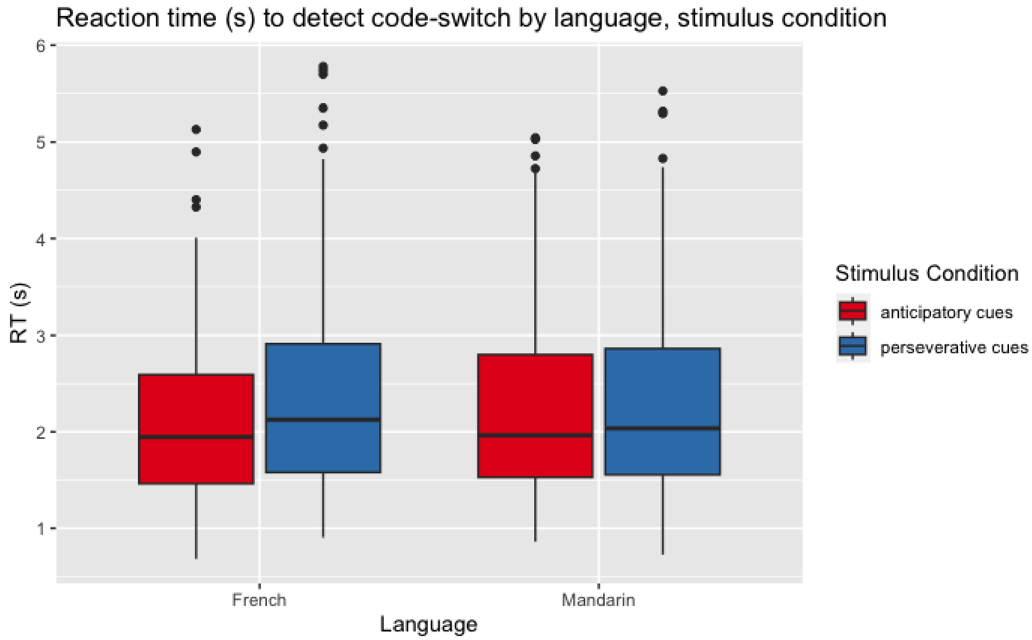


Figure 9: Experiment 2: Reaction time (s) to detect code-switch by age, language, stimulus condition

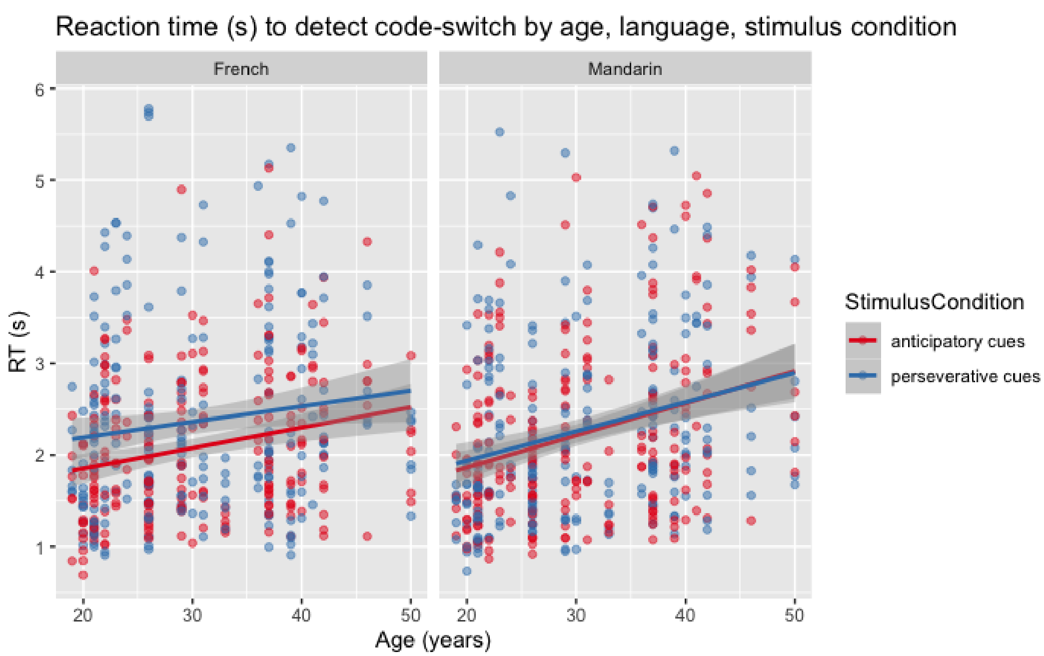
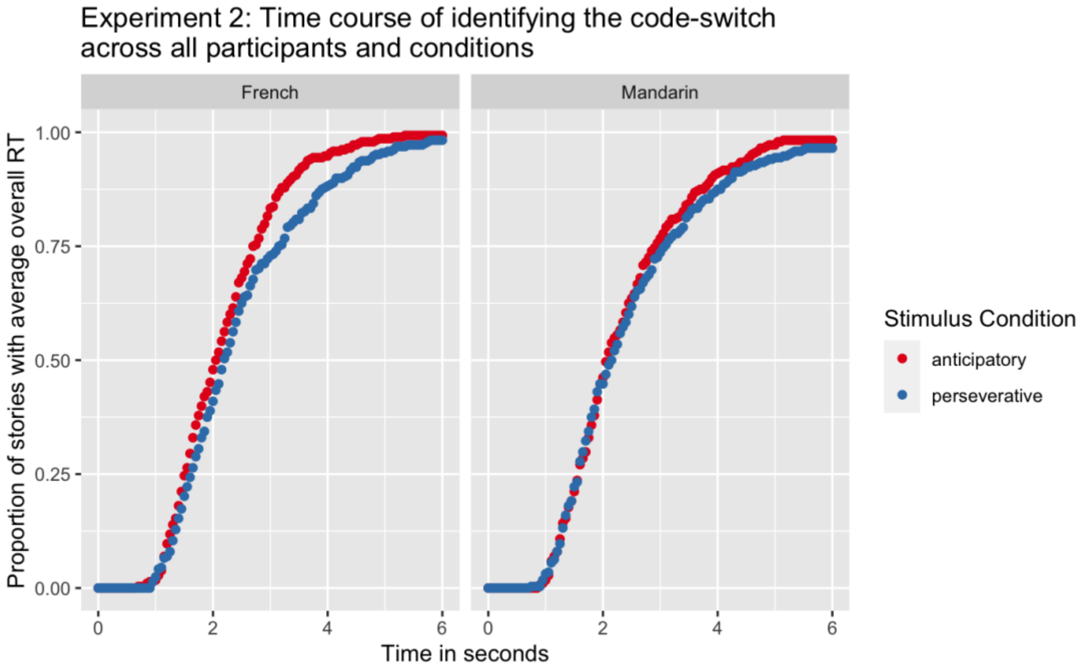


Figure 10: Experiment 2: Time course of identifying the code-switch across all participants and conditions

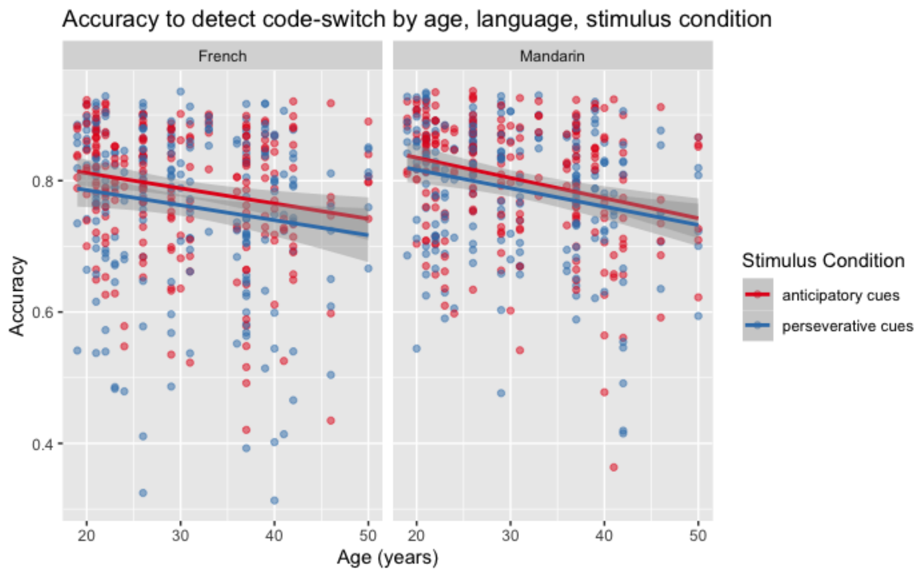


Statistical analysis of accuracy:

The maximal model that converged included random intercepts of participant and passage, random slopes of language within participant and random slopes of condition and language within passage. This model showed a significant effect of language such that participants identified Mandarin more accurately than French ($t(14.63) = -2.24$, $\beta = -0.064$, $p < 0.05$), of condition such that participants were more accurate for segments that had anticipatory cues than for segments that had perseverative cues ($t(11.0) = 2.555$, $\beta = 0.058$, $p < 0.05$), and of age such that older participants were less accurate ($t(34.55) = -3.017$, $\beta = -0.018$, $p < 0.01$). No interactions were significant in this model, including the interaction between language and condition which was significant in the model of reaction time.

A simplified version of this model did not significantly differ from the full model, $\chi^2(7) = 5.93$, $p = 0.55$. This simplified model continued to show significant effects of language (t(14.93) = -2.232, β = -0.064, $p < 0.05$), condition (t(11.0) = 2.56, β = 0.058, $p < 0.05$), and age (t(37.62) = -2.917, β = -0.017, $p < 0.05$).

Figure 11: Experiment 2: Accuracy to detect a code-switch by language, condition, age



<u>Discussion</u>

In experiment 2 I replicated the result from experiment 1 showing that adults can identify which language they are hearing when presented with a speaker who regularly switches between French and Mandarin. Further, in both studies I showed that Spanish-English bilingual listeners are sensitive to the small ways that a speaker's production of one or both languages may change immediately surrounding the point of a code-switch. Specifically, experiment 2 suggests that the presence of such changes prior to the point of a code-switch is more beneficial for speed of language identification than the presence of such changes after the point of a code-switch.

There is an interaction between language and condition on reaction times, such that listeners were faster to identify French when provided with anticipatory cues rather than perseverative cues, but that there was no evidence of an effect of condition for switches into Mandarin. However, there was no overall effect of language; when averaging across conditions, reaction time was 2.23 seconds to identify French and 2.25 seconds to identify Mandarin. Instead, this interaction resulted in *faster* identification of French when provided with anticipatory cues compared to overall speed to identify Mandarin, and *slower* identification of French when provided with perseverative cues compared to overall speed to identify Mandarin.

Presence of acoustic cues and effect of language choice

This Mandarin-to-French switch direction is comparable to the English-to-Spanish switch direction that was assessed in a comprehension paradigm in Fricke et al. (2016). Mandarin, like English, uses relatively long VOTs for voiceless aspirated stops, while Spanish, like French, uses relatively short VOTs for voiceless stops. English VOTs have been reliably found to be produced slightly shorter, and therefore more Spanish-like, close to the point of a code-switch between English and Spanish. Similar patterns also occur in English-Greek code-switching (Antoniou, et al., 2011). Therefore, we expect that these sub-phonemic changes in VOT would also be present in Mandarin-French code-switching, with Mandarin VOTs produced slightly shorter close to French. Fricke et al. (2016) showed that participants were more easily able to comprehend a switch into Spanish when the preceding English included these naturalistic acoustic cues to code-switching. Experiment 2 mirrors that finding with the Mandarin-to-French switches. However, Fricke et al. did not assess the other direction, Spanish-into-English code-switches. Production research is mixed regarding the state of phonetic transfer on voice onset time in these short-lag languages like Spanish and French. While some work shows evidence for VOT changes in both directions, others find these effects to be asymmetrical. From a typological perspective, it is possible that VOTs in the short-lag language simply have less room to move (Bullock et al., 2006), because either those long VOTs are not regularly used in the language in any phoneme, or the relatively small ~0-30ms range does not allow for much variability. Therefore, if VOT is a primary anticipatory acoustic cue that listeners are using in the Mandarin-to-French direction, it is less likely that we would see similar effects in the French-to-Mandarin direction as French VOTs may not lengthen in anticipation of upcoming Mandarin to the same extent that we expect Mandarin VOTs decrease in anticipation of upcoming French.

French and Mandarin were also selected for the potential presence of changes in intonation, as discussed in Shen et al. (2020). They found that English-Mandarin bilinguals had more difficulty comprehending the word directly following an English-to-Mandarin code-switch when the English did not contain natural acoustic cues to code-switching as compared to a fully naturalistic sentence. A subsequent acoustic analysis, however, was unclear with respect to whether intonation patterns preceding the code-switch are tone-specific in a way that would suggest preparation for the upcoming target word in Mandarin and might therefore help the listener to predict a word using that Mandarin tone. This study, therefore, presents another approach to studying a similar pattern, though using French instead of English. While I did not find strong evidence suggesting that French intonation immediately preceding Mandarin had properties that made it easier for listeners to predict and identify Mandarin in experiment 2, that possibility cannot be completely ruled out without testing additional participant populations. That is, Spanish-English bilingual participants were likely familiar with acoustic changes in the dimension of voice onset time, but not in lexical tone as they do not speak a tone language. If a listener's familiarity with the acoustic dimensions that may be relevant in code-switching for a given pair of languages determines their sensitivity to these natural cues to code-switching, then we would expect speakers who know at least one tone language, such as Vietnamese-English or Igbo-English bilinguals, to benefit from anticipatory cues to code-switching driven by lexical tone (should such cues exist). Bilinguals without familiarity with lexical tone though, such as the Spanish-English bilinguals who participated in experiment 2, might not have the language knowledge or experience necessary for such a benefit.

Two additional possible explanations regarding the existence and utility of anticipatory and perseverative cues to code-switching in this French-to-Mandarin direction remain. First, it could be that cues in both positions are present and equally strong, such that comparing the two conditions effectively cancels out their effects. Similarly, it could be that both types of cues are weak or absent in code-switching from French into Mandarin. Acoustic analysis designed to assess the detailed acoustic-phonetic features of French and Mandarin in each of these conditions may be able to address this question, by measuring voice onset times in both languages and testing whether those voice onset times are affected by an upcoming (or immediately preceding) language switch in the natural code-switching recordings. In the French-to-Mandarin direction specifically, acoustic analysis should also assess French intonation contours preceding a language switch for each individual Mandarin tone that begins the switched-into segment, in an analysis patterned after that in Shen et al. (2020).

Issues of identifying the appropriate 'baseline' condition

As in experiment 1, it is difficult with the current study design to assess whether these results should be interpreted as facilitation due to the presence of anticipatory phonetic transfer and/or lack of perseverative phonetic transfer (for French), or as increased difficulty due to the presence of perseverative phonetic transfer and/or lack of anticipatory phonetic transfer. That is, we do not know if the relative ease in identifying French in the anticipatory-only condition should be attributed to the presence of preceding anticipatory phonetic cues (produced in Mandarin), or the lack of perseverative effects of Mandarin affecting production of French. We also do not know which, if any, condition should be interpreted as the "baseline" in this experiment, especially because neither condition is truly naturalistic. Both code-switch conditions contain one segment which was produced with natural code-switching and one that was produced in a unilingual passage. The anticipatory-only condition is, however, somewhat more natural than the perseverative-only condition because all of the information heard prior to the code-switch, which has the potential to bias or influence speech perception leading up to that point, is naturalistic for a code-switching context. The information heard prior to the code-switch for the perseverative-only condition would be naturalistic for a unilingual context, but participants know that they are in fact in a code-switching context.

Further, study 2 resulted in overall shorter reaction times and higher accuracy than study 1, though by a very small margin. Overall bilingual participants in study 1 correctly identified the language in 2.39 seconds and for 76.5% of the time, while they did so in 2.19 seconds and for 79.11% of the time in study 2. As these were entirely independent groups of participants completing unique tasks, their reaction times should not be directly compared. However, the groups are demographically balanced, and the bilingual and monolingual groups in study 1 do *not* differ in their overall reaction time or accuracy to identify the correct language in that task. Therefore, one possible interpretation regarding the cross-study comparison is that identifying code-switches when only anticipatory or perseverative cross-language phonetic transfer is present is legitimately easier than identifying code-switches when this phonetic transfer is either entirely present or entirely absent. If such a pattern is true, it may suggest that naturally-occurring perseverative effects of code-switching actually make it harder for listeners to identify the language being spoken. Both anticipatory and perseverative transfer may serve to effectively "blend" the phonetic detail of the two languages and make them sound slightly more similar. While experiment 2 (and previous research) suggests that this coarticulatory pattern may be useful when it only precedes the language switch, perhaps perseverating phonological activation of the now-unused language on the part of the speaker makes the languages sound so similar that it is harder to identify that a switch occurred. Listeners may need to accumulate more information to identify that second language, which would result in longer reaction times when perseveration occurs.

Rather than accumulating tokens and comparing them to an existing distribution of speech sounds in each language, listeners in the present studies may have instead approached the task by listening for speech sounds or words that sound obviously like one or the other language, for example by focusing on segments that are used only in French or Mandarin. Perseverative phonetic transfer may cause segments appearing shortly after a code-switch to sound less language-extreme than they do when produced in a unilingual context, such that more time elapses before a segment occurs which triggers a participant to move their mouse to the other box. This "triggering" hypothesis is somewhat specific to the current situation where listeners are unfamiliar with the two languages, such that words and phonemes may need to be particularly extreme and obvious to signal a language change to the listener. In previous studies investigating sensitivity to natural phonetic cues to code-switching, and in most real-life situations involving code-switching, the listeners are fluent speakers of both relevant languages and primarily aim to comprehend the speech rather than to consciously identify the language. Identifying the language is rarely strictly necessary for speech comprehension, and indeed some examinations of code-switching comprehension using ERPs shows that in both reading and listening paradigms the neural signatures of language identity and lexical meaning interpretation diverge, suggesting that the two processes are independent (Blackburn & Wicha, 2022; Ng & Wicha, 2013; Yacovone et al., 2021).

Future research could further investigate these two hypotheses by experimentally manipulating the location and phonetic properties of words and phonemes that listeners are most likely to be able to identify as being generated by a particular language. The triggering hypothesis would suggest that participants identify a code-switch immediately after one of these tokens occurs in a language-extreme form. In contrast, the accumulating hypothesis predicts that all information which both precedes and follows a code-switch point is used, and the distribution of these surface forms is compared to that of unilingual productions; a ‘trigger’ word or phone alone is not enough to signal a switch to the other language.

Experiments 1 & 2: Interim discussion

Comparing results from Experiments 1 & 2

The results here suggest that Spanish-English bilinguals who have experience listening to code-switching in their two languages are sensitive to natural cues to code-switching in two languages with which they are not familiar (French and Mandarin Chinese). However, bilinguals crucially were not overall better than age-matched monolinguals at identifying which of two unfamiliar languages is currently being played. This finding leads to two related conclusions regarding the interaction between experience and speech perception.

First, the difference in observed patterns based on experience knowing multiple languages and listening to code-switching suggests that listeners are not universally sensitive to sub-phonemic shifts in phoneme production within a speaker and capable of abstracting these shifts to language categorization, but rather that such sensitivity is a result of lived experience with bilingualism and code-switching. The results from previous research that tested the effects of natural cues to code-switching on comprehension while Spanish-English bilinguals listened to Spanish-English code-switching may have been driven by the fact that the cues tested were familiar to the listeners. Corpus studies of Spanish and English suggest that the dimensions of pitch, VOT, and speech rate change in the presence of a code-switch and could signal such a switch. Perceptual studies investigating the comprehension of these two languages therefore would include these same cues. Thus, previous studies tested fluent bilinguals’ perception of the exact cues that they themselves may have produced and heard regularly. The finding here that a sample of bilinguals from that same population used similar cues in to discriminate between a pair of relatively unfamiliar languages suggests that this prior experience generalizes to impact how impact listeners perceive even novel languages.

Several possibilities regarding the nature of this generalization and how it affects bilinguals' representations of unfamiliar languages exist. It could be that being a bilingual [particularly one who code-switches or hears code-switching regularly] and managing multiple linguistic representations including phonological systems *increases individuals' metalinguistic awareness and phonological sensitivity, making them more attuned to various forms of allophonic variation both within a language and between languages*. Using this increased metalinguistic awareness and phonological sensitivity, bilinguals may be better able to track the distribution of acoustic realization of phonemes in each language over time and to notice when tokens are produced outside their expected distribution in a way that could signal code-switching. If this explanation is the case, we would expect bilinguals to be better than monolinguals at distinguishing between two speech sounds which are allophonic in all of their languages but phonemic in some other language. While infants in bilingual environments are sensitive to non-native phoneme contrasts later in development than their monolingual peers (i.e. Singh et al., 2017), very little is known about how novel phoneme discrimination differs between monolingual and bilingual *adults*. However, one study found that if anything, monolingual adults were better than bilingual adults at an ABX phoneme identification task using a contrast that was novel to all participants (the Bengali dental-retroflex contrast; Burfin et al., 2014). Thinking beyond phoneme discrimination and considering the relationship between acoustic form and representation, however, bilinguals may be better able to map a known phoneme to a novel surface representation (Spinu et al., 2018) and to adapt to characteristics of an arbitrary non-native accent (Weber et al., 2014) than monolinguals. Therefore, it is possible that bilinguals, as a by-product of using multiple phonetic systems, have somewhat more flexible phonetic representations than monolinguals do.

Another, not necessarily mutually exclusive, possibility is that bilinguals are *specifically attuned to distributional shifts in phonetic realizations that are similar to patterns with which they are familiar in their own languages***.** Under this theory, being bilingual (and specifically hearing code-switching frequently) led to the observed effect of condition in experiment 1 not necessarily because bilinguals manage representations of two interactive phonetic and phonological systems, but because they frequently hear similar patterns. This proposal is in line with the production-distribution-comprehension model (MacDonald, 2013), especially as outlined in its relationship to phonetics and bilingual code-switching (Fricke et al., 2016).

One method of assessing these possibilities may be to evaluate whether participants' performance changes over time, throughout trials of the experiment. If the observed effects are simply due to pre-existing expectations about the phonetic properties of code-switching, then we would expect to see relatively consistent effects of condition across all trials (though reaction times may generally improve as participants become more familiar to the task and able to differentiate between French and Mandarin). If, however, some participants are simply more attuned to allophonic variability and able to find distributional phonetic patterns, then the effect of condition may increase trial-to-trial.

Second, I found evidence that anticipatory cues are particularly helpful in listeners' ability to detect a language switch and that they outweigh perseverative cues (at least for phonetic changes and dimensions which are likely to be more familiar to listeners and acoustically salient, i.e. a language with generally long VOTs produced with slightly shorter VOTs preceding a switch into a language with relatively short VOTs). This finding that anticipatory cues dominate suggests that this natural switching may aid *prediction* of an upcoming switch rather than *identification* of the newly-switched-into language. Previous studies that found similar effects of natural anticipatory cues (Fricke et al., 2016; Shen et al.,

2020) did not require conscious language identification, as they studied word comprehension immediately following a switch in fluent bilinguals. Measurements also occurred over a shorter time period, with participants fixating more on the target word than an interlingual distractor within 750 ms post-onset of the target word. Meanwhile, in this study it took participants approximately 2-2.5 seconds to identify the current language, with some participants taking much longer or even not recognizing the language at all on individual switches. Due to this delay, the exact cues that listeners are sensitive to, or the amount of time it takes them to use a cue to recognize a language switch after it has been produced, is not clear. It appears that the relevant section of input is separated from the response that it affects by several seconds, which indicates that distributional shifts in speech may influence perception in ways that are not necessarily immediate[7].

As previously discussed, the current results are compatible with both a model of "accumulation," whereby participants are constantly comparing the distribution of surface forms to that of underlying representations and waiting until the bulk of evidence has reached some threshold, and "triggering," whereby listeners respond immediately following some word

---

[7] Another possibility is that perseverative cues are helpful, but because they occur later in time than anticipatory cues, it takes longer for them to functionally impact reaction time in response to a code-switch. That is, in experiment 2, trials that could only contain anticipatory cues were directly compared to those that could only contain perseverative cues, and reaction times were faster in the anticipatory condition than the perseverative condition, indicating that participants were faster to identify a language switch when sub-phonemic acoustic cues to that code-switch could have been available before it occurred, than when they only could have been available after it occurred (during the time period of response). This finding does not rule out the possibility that reaction times in the perseverative condition were faster than they would have been without such perseverative cues. Testing such an effect would require comparing the perseverative-only condition to trials without any acoustic cues to code-switching. While this fully-unnatural condition was not present in experiment 2, it was in experiment 1. In the perseverative condition in experiment 2, participants responded in 2.37 seconds for switches into French and 2.27 seconds for Mandarin. In the unnatural condition in experiment 1, bilingual participants responded in an average of 2.44 seconds (no effect of language). This difference suggests that the perseverative-only condition may have provided some benefit to participants in their reaction times; however, as this is a between-subjects comparison, the results are tentative and could have been driven by overall differences in speed of reaction between the participant samples in experiment 1 and experiment 2. Future research that includes all four conditions (unnatural, anticipatory only, perseverative only, fully naturalistic) within the same participants will be necessary to properly address this question.

or token that obviously belongs to a particular language. Another possibility that combines both accounts is that some aspect of these anticipatory changes may prime participants to listen for an upcoming signal indicating language identity, and that listeners are then better prepared to both perceive and respond to that language-specific cue when it arrives in the seconds immediately following a language switch (or that the lack of anticipatory changes make it harder for listeners to perceive and respond to a language-specific signal due to being actively unprepared).

To investigate the specific role of these anticipatory distributional shifts in speech production and how they influence a listener's perception, future work should assess how listeners respond when phonetic cues would predict a code-switch but one does not occur. Using the stimuli from the current studies, for example, I could create a passage that has one Mandarin segment with natural switches to code-switching which is followed by another segment in Mandarin (which could be either natural or unnatural). The current study shows that listeners are better at identifying French following this initial $\text{Mandarin}_{\text{natural}}$ segment than following $\text{Mandarin}_{\text{unnatural}}$ segments, and that Mandarin produced with natural code-switches therefore contains useful anticipatory cues to French. If listeners false alarm and incorrectly indicate that French occurred following this natural Mandarin segment (rather than continuing to indicate that Mandarin is playing) more often than they do following an unnatural Mandarin segment, it would suggest that hearing the anticipatory phonetic cues leads listeners to actively predict (or think that they are hearing) French. If, however, false alarm rates to French do not differ for $\text{Mandarin}_{\text{unnatural}}$-Mandarin and $\text{Mandarin}_{\text{natural}}$-Mandarin switches, it would suggest that hearing the anticipatory phonetic cues serves a role in orienting listeners' attention to the acoustic features that may be relevant in correctly identifying an upcoming language.

As the current experiments do not find consistent effects in the accuracy dependent measure, and a design based on rates of false alarms necessarily depend on accuracy, it is not clear that such effects would be detectable. However, whereas the current accuracy measure is calculated based on the proportion of time that the participant spends correctly identifying a given language (rather than identifying the wrong language or neither language), this proposed false alarm measure would look at *whether* participants move their mouse into the incorrect [French] box in the first 5 seconds of the second consecutive Mandarin segment, in addition to the proportion of time they indicate the incorrect language. This binary measure of accuracy may be better able to detect instances when participants think they hear a code-switch but then quickly change their minds and move their cursor back to the correct box.

A study addressing this same question would also be possible using a similar word comprehension visual world paradigm as Fricke, et al. (2016) and Shen et al. (2020) and testing bilingual participants in the two languages which they speak. It would include unilingual trials in which the sentence frame is spliced from an originally code-switched recording, with phonological cohort competitors in both languages. If participants look to the phonological cohort competitor in the language that differs from the sentence frame more than the phonological cohort competitor in the language that matches the sentence frame, it would suggest that they predicted an upcoming language switch.

Further, the effect observed in experiment 1, that only those bilinguals with more experience listening to code-switching showed an effect of stimulus condition, raises the more foundational question of whether the other participants did not *perceive* a difference between the passages that did and did not contain naturalistic acoustic cues to code-switching, or if they perceived the difference but could not use it to *detect* a code-switch. To address these different explanations, a future study could specifically ask participants to discriminate between short (2-5 second-long) segments that have identical content but derive from either the unnatural or the naturalistic stories. If bilinguals (especially those with more code-switching experience) are better at identifying whether two content-identical passages are acoustically-identical than monolinguals or bilinguals with minimal code-switching experience are, it would suggest that the benefit is in low-level perception. If, however, there is not an effect of language experience on accuracy in such judgments, then we could conclude that previous linguistic experience instead affects listeners' ability to make higher-level predictions about the relationship between phonetic detail and code-switching.

Limitations and future directions

The present experiment only tested perception of one language pair (Mandarin-French) spoken by one speaker (YM). The results show promising evidence that Spanish-English bilinguals perceive sub-phonemic anticipatory cues to code-switching in these two languages, and use these cues to predict an upcoming language switch. However, the generalizability of these results to additional speakers and languages is unclear. It is possible that speakers differ in the degree to which they produce such phonetic convergence, and that these differences are driven both by language dominance and idiosyncratic production behaviors.

Further, the code-switching behaviors of specific bilingual communities vary greatly and may be influenced by historic cultural migration patterns, including the societal power and integration of speakers of those languages in a given region (Nilep, 2006), though more research on this topic is needed. In North America, for example, code-switching is culturally important for many Spanish-English bilinguals in the United States as a way to emphasize their Hispanic identity, while French-English bilingual speakers in Canada may not use code-switching for the same purpose (Kircher, 2014).

Beyond rates of code-switching, perhaps individuals who belong to bilingual communities for which code-switching is a common behavior are more likely to produce this phonetic convergence because they are familiar with the coarticulatory patterns associated with code-switching. To the author's knowledge, no research regarding the frequency or properties of code-switching in France-based Chinese communities exists, and therefore it is impossible to truly assess how this linguistic practice may differ in YM's community compared to other bilingual settings. However, work comparing code-switching in a variety of Spanish-English communities, and in individuals with a range of proficiency and relative dominance levels, suggests that such community- and identity- based factors may influence production of code-switching at the acoustic-phonetic level (Balukas & Koops, 2015).

Further, it is possible that because two languages which have VOT distributions approximately corresponding to those of English and Spanish were chosen, bilingual adults' abilities to use acoustic cues to code-switching were maximized. If, instead, no available cues to code-switching mirrored the cues in Spanish-English code-switching, it would have been possible to test the degree of generalization that occurs between the native languages and a pair of novel languages. As previously discussed, if bilingualism leads to a generally increased phonological sensitivity, then we would expect similar patterns to appear for any language pair, including those for which the acoustic dimensions which show the greatest change due to code-switching differ from the dimensions and directions of change in the language pair that a particular listener does know. However, given the lack of an effect in the French-to-Mandarin direction in experiment 2, if there are anticipatory phonetic changes in those types of code-switches, it may suggest that bilinguals do not benefit from all acoustic cues.

Due to these possible sources of variance in the phonetic detail of code-switching (individual, community, language), the use of a single speaker of a single language pair limits the generalizability of the reported findings and a broader understanding of how listeners perceive code-switching in unfamiliar languages. However, the initial results reported here, that (at least some) listeners *are* sensitive to naturally-occurring acoustic cues to code-switching and are better able to detect a language switch when such cues are present before the switch occurs, provide important evidence regarding listeners' sensitivity to sub-phonemic detail when listening to speech that is unintelligible to them. The ability to subconsciously detect a multi-dimensional distributional shift in the acoustic and phonetic properties of speech is not conditional on speech *comprehension*.

Without an acoustic analysis, it is impossible to determine exactly what cues are present in the stimuli, or when those cues occurred relative to a code-switch. As discussed, an analysis of the voice onset times of voiceless stops in French and Mandarin in both the natural code-switching recordings and unilingual recordings will be necessary to determine whether Mandarin VOTs did in fact decrease preceding a code-switch into French, as predicted based on both previous Spanish-English corpus analyses and the pattern of behavioral results seen in experiments 1 and 2. Such an analysis can also assess whether French VOTs were impacted by the presence of and proximity to Mandarin.

An important component of this acoustic analysis will be to uncover the time period over which such changes apply. That is, in addition to determining whether they are anticipatory or perseverative in nature, such analyses should assess whether these acoustic cues are linear throughout a segment or if this convergence only extends in a small range. That is, are Mandarin VOTs gradually shorter approaching a code-switch into French, are they relatively constant up to a certain point before linearly decreasing when approaching that code-switch, is there a point of discontinuity at which VOTs jump from one relatively constant value to a shorter, relatively constant value, etc. A piecewise regression designed to find the elbow and the coefficient of regression at each time range will help to answer this question.

This acoustic analysis should assess dimensions beyond simply voice onset time, such as pitch contours. Shen et al. (2020) analyzed English pitch contours preceding a code-switch into Mandarin and found some evidence for tone-specific coarticulation such that the speaker may have been preparing to produce a particular tone in Mandarin at that code-switch point. A similar analysis here, which measures French pitch contours preceding a code-switch point in the unilingual recordings and the natural code-switching recordings, would allow us to test the robustness of such an effect. If there are anticipatory cues to code-switching in pitch contour, then we would expect tone-specific co-articulation in the natural code-switching recordings and not in the unilingual recordings.

Determining the exact acoustic cues present in the stimuli may further help to clarify the relationship between participants' language backgrounds (and the cues to code-switching that may be familiar to them) and their sensitivity to such cues in unfamiliar languages. As discussed, it could be that listeners particularly benefitted from anticipatory cues in experiment 2 because such cues were present to a greater degree and thus more available than perseverative cues. However, as established in the production-distribution-comprehension account (MacDonald, 2013), listeners may be more sensitive to the types of cues that are more familiar to them. Such increased sensitivity could apply both to acoustic dimension (i.e. VOT), but could also be true of temporal position, such that listeners who are accustomed to anticipatory phonetic convergence then expect some cue to a code-switch before that switch occurs in unfamiliar language pairs. If an acoustic analysis finds evidence of perseverative phonetic convergence to the same degree as anticipatory convergence, it may be that listeners relatively undervalued such a cue due to its lack of prominence in their existing mental models.

Chapter 4: Perception approach: children and infants (Experiments 3-4)

Experiment 3: Children

Introduction

As previously discussed, children, like adults, show some ability to distinguish between two languages that they do not know, though perhaps with lower accuracy. Preschoolers (4-5 year-olds), school-age children (7-8 year-old) and adults can identify whether two short excerpts spoken by different speakers are derived from the same language or different languages with above-chance accuracy (Stockmal, 1995; Stockmal et al., 1994). However, preschoolers struggled to identify two samples as being "the same," especially when the information provided to make that judgment was only 2 seconds in either language. Longer excerpts improved performance, though older listeners were more accurate and appeared to use more detailed information including segmental cues and knowledge of a small number of individual words whereas younger listeners relied more on prosodic characteristics of each sample (Stockmal, 1995).

The task is noticeably harder when, as in the present study, both languages are produced by the same speaker. Bond & Stockmal (2005) presented English monolingual children with bilingual speakers producing excerpts in two languages (Arabic-French, German-Hebrew, Japanese-Korean, French-Mbawa) and asked the children whether the two sentences spoken by that person were in the same language or different languages. Both 4-5 year-olds and 7-8 year-olds struggled, though some 7-8 year olds did perform the task reliably. In a slightly easier version of the same task using only the Japanese-Korean and French-Mbawa pairs, 4-5 year old children continued to perform at chance accuracy (Marks et al., 2009).

However, 7-8 year-olds were able to reliably indicate whether the two sentences were from the same or different language. Overall results between English monolinguals in the United States, Spanish monolinguals in Spain, and Spanish-Basque bilinguals in Spain did not differ (Marks et al., 2009), which indicates that overall language and cognitive development may impact this meta-linguistic awareness, rather than experience with multiple languages per se.

However, as previously discussed, task-internal familiarization could impact performance by biasing the features to which children respond. When familiarized with an unknown language that is irrelevant to the discrimination task, 5-6 year-old monolingual children were actually *better* at discriminating unknown languages (Italian-Spanish or Mandarin-Spanish) compared to when they were familiarized with one of the two target languages or with English (Potter & Saffran, 2015). Children performed at above-chance accuracy in all conditions, even in Potter & Saffran's ABX task which required a greater degree of conceptual categorization than the same-different tasks utilized by Bond, Stockmal, Marks, and colleagues. This pattern of results across these studies suggests that 1) children can attend to segmental information in short segments of novel languages and use that information to create conceptual categories of multiple linguistic codes and that 2) age, linguistic experience, and task demands may influence the specific segmental features to which children attend when listening to two unknown languages. In experiment 3 I aim to extend these findings by asking children to continually categorize two languages, rather than using an offline judgment task.

Experiment 1 tested monolingual and bilingual adults' ability to distinguish between Mandarin and French, two languages with which they were unfamiliar, and crucially to use sub-phonemic acoustic cues to code-switching to distinguish between the languages. Though all participants reported minimal or no exposure to both French and Mandarin, it would be very unusual if they did not have at least some sense of what each language sounds like, especially as adults living in the United States. This pre-existing exposure, however minor, could have impacted participants' performance in the task and particularly their sensitivity to minor distributional shifts in the phonetic properties of the two languages. That is, in order to respond more quickly to the naturalistic as compared to the unnatural condition, participants may need to have a rudimentary phonological representation of each of the languages, such that they not only can identify which language is being spoken at a particular time, but can compare the current input to that of their representation of allowable variability for each language. In the current study, I address this concern by assessing the cues that children use to detect when the language switches between French and Mandarin.

In a pilot study, I established that English monolingual children can complete the same task as adults and identify code-switches between English and Mandarin. Fourteen children between the ages of 5.25 and 8.08 years (M = 6.43) listened to a woman tell six stories while switching between English and Mandarin Chinese. All stories were recorded in the unilingual condition and spliced to create code-switching recordings. This pilot study therefore did not allow for a comparison between the naturalistic and unnatural conditions or to separate effects of sound patterns vs. semantic effects, though it did show that children of the same age can comprehend and perform the task.

Though children listened to six stories in this pilot study, I discovered later that one of the stories was spliced incorrectly due to experimenter error. Results were therefore analyzed for only five stories. Six of the fourteen children identified the correct language for at least half of the story for all five analyzed stories; the other eight did so for four of the five analyzed stories. In considering only those 62 stories in which children responded correctly at least half of the time, I then assessed accuracy and reaction times for English and Mandarin across each story, as in experiment 1. Children identified Mandarin correctly an average of 71.53% of the time (min = 58.72%, max = 82.84%). They identified English correctly an average of 72.9% of the time (min = 58.8% of the time, max = 84.7% of the time). On average it took children

2.18 seconds to identify Mandarin (range: 1.26 – 3.87 seconds) and 1.84 seconds to identify English (range: 1.06 – 2.73 seconds). In looking at the individual results, five of the fourteen children were both faster and more accurate at identifying English than Mandarin, four were faster to identify English but more accurate at identifying Mandarin, two were more accurate at identifying English but faster to identify Mandarin, and three were both faster and more accurate to identify Mandarin.

Based on the results from the pilot study, we can conclude that 5- to 8-year-old children can move a cursor on a computer screen while listening to a story that switches between two languages, and use that cursor movement to continually indicate which language they are hearing. However, they were slightly faster to identify their native (and thus, a familiar) language than a language they did not know. The exact method children were using to perform the task is therefore unclear. One possibly is that they were relying on comprehension of their native language, and essentially using a heuristic where the language was English if they understood the speech and Mandarin if they did not. Such a method would not require children to have any representation of Mandarin, and predicts that they would not succeed at the task if faced with two unknown languages. However, an alternative possibility is that children were using phonological representations of both English and Mandarin to some extent, rather than solely lexical knowledge. They may have been faster to identify English than Mandarin because those phonological representations were more well-formed, but still have been positively identifying Mandarin rather than concluding that it must have been the language through a process of elimination. The first possibility, with an “I understand this” heuristic, does not require deep meta-linguistic awareness of what it means for multiple languages to exist or how two languages may differ, though the second does require some meta-linguistic awareness. In experiment 3 I test these hypotheses by asking children of the same age to perform the same task, but to instead listen to two unknown languages. Experiment 3 also tested whether children are sensitive to natural acoustic cues to code-switching, as a comparison to experiment 1.

Procedure

Children aged five to eight years old were recruited through Children Helping Science powered by Lookit (Scott & Schulz, 2017). Any family who had an active account on Children Helping Science with at least one child between the ages of 5;0 and 8;12 who was only exposed to English could see the study and choose to participate; a random selection of these families also received notification emails encouraging them to participate. Participants and their parents or legal guardians provided a recorded statement of informed consent and their verbal assent on the Lookit platform, and then watched a video of the experimenter introducing them to the task. Parents of participants also completed a demographic and language background survey for their children, designed to ensure that the children were in fact English monolinguals and had little if any exposure to either French or Mandarin.

Following consent and the questionnaire, participants were then redirected to the Cognition.run platform to participate in the same task as experiment 1, with a few minor adjustments for the participants' age. While each adult listened to 12 passages (6 in each condition), children listened to a total of 6 passages (3 in each condition), as their attention spans were not expected to last for all 12 passages. Children therefore only heard one version of each of the six stories (either the Mandarin-first order or the French-first order; mapping between language order and condition was counterbalanced across children). However, the passages themselves were identical across the two age groups. Children also saw additional visual indicators to help them identify which side of the box was associated with each language (the Eiffel Tower and the Great Wall of China), in case their reading abilities would not allow them to quickly glance at the screen and decode the labels[8]. Parents supervising their children were told to read the instructions and answer clarification questions, but to avoid offering feedback or support on children's selections during the study.

Several changes were also made to the analysis pipeline for children compared to adults, to reflect age-based behavior and performance effects. Observation of the children's responses shows consistent behaviors that did not appear in the adult data. Specifically, the children were more likely to rapidly move their mouse between the two boxes, with some children moving their mouse between boxes over 100 times in a story lasting 90 seconds. No adult did this. Children also sometimes left their mouse in one box for an entire story, or did not select either language for more than half of a story. Again, these behaviors were very rare for adults, especially in comparison with the children.

Children's responses were scored in three ways: accuracy time, correct rejection time, and correct moves. Accuracy was scored the same way for the children as it was for the adults: when a given language was playing, what proportion of the time was the participant's mouse in the box corresponding to that language? Accuracy is thus a durational measure, operationalized as a proportion. Correct rejection can be seen as the opposite of accuracy: when a given language was *not* playing, what proportion of the time was the participant's mouse *not* in the box corresponding with that language? That is, how often did they reject a given language when it was the wrong answer? Note that a participant's correct rejection score for one language

---

[8] I acknowledge that these images are specific to individual countries, and thus do not reflect the broad scope of speakers of Mandarin or French across the world. However, they are widely-recognized, and reflect the language varieties spoken by YM as he grew up in China and lived in France as an adult, including at the time of recording. A potentially more important caveat is that these images were not necessarily recognizable or useful to the children who participated in the study. Five-year-olds are unlikely to have learned about either the Great Wall of China or the Eiffel Tower in their standard curriculum at school, though individual children may have encountered these cultural artifacts through conversations or activities with their families. The usefulness of adding these images to the screen is therefore debatable, as the children most likely to know what they mean enough to obtain some benefit would be the stronger readers who did not require a non-linguistic label for each box. However, the images are unlikely to have caused any harm or added to the difficulty of the task, for either the new readers or the strong readers.

is necessarily greater than or equal to their accuracy score for the other language, as time when the participant's mouse is outside the two boxes is counted positively toward the correct rejection score for the language that is not playing and negatively toward the accuracy score for the language that is playing. As children (unlike most adults) frequently moved their mouse outside of both boxes, their correct rejection scores are higher than their accuracy scores in the opposite language. The table below shows how each possible combination of ground truth and participant response counted toward the accuracy and correct rejection measures for each language.

Table 3: Experiment 3: Categorization of each response type

	Mouse in Mandarin	*Mouse in French*	*Mouse outside both boxes*
Ground truth: Mandarin playing	Accurate for Mandarin	NOT accurate for Mandarin	NOT accurate for Mandarin
	Correct Rejection for French	NOT Correct Rejection for French	Correct Rejection for French
Ground truth: French playing	NOT accurate for French	Accurate for French	NOT Accurate for French
	NOT Correct Rejection for Mandarin	Correct Rejection for Mandarin	Correct Rejection for Mandarin

To calculate the correct moves score, I tallied the total number of times the participant moved their mouse into the box associated with each language, and then determined whether each of those moves was correct (that language was playing) or incorrect (that language was not playing). The correct moves score is simply the proportion of moves into each language which were correct. In this way, the correct moves score is normalized for each passage; a passage in which the child moved their mouse 100 times is weighted equally to a passage in which a child moved their mouse only twice. Note that because children could have moved their cursor outside of both boxes and then back into one of the two boxes, the number of moves into one language is not necessarily related to the number of moves into the other language.

A passage was excluded from further analysis if the participant received a score less than 0.10 in either language for the accuracy time dependent measure. Any participant for whom 3 or more passages was excluded was not included in the final analysis. Whereas the accuracy cutoff guidelines for the adult participants were designed to ensure that the adults understood the task and generally knew which language was playing, the accuracy cutoffs for the children were designed to ensure that they participated sufficiently in the activity by moving their mouse enough to correctly identify each language at least 10% of the time. Half of the child participants did not exceed 50% (chance) accuracy in any of the six stories. On average, participants performed above 50% accuracy in both languages for 0.88 out of 6 stories, with a maximum of 5 (n = 1). Using the same accuracy guidelines for both adults and children may have required testing over 400 children to reach a sample size of 40, leading to a deeply unusual sample of children and an inability to generalize results to a broader population.

It therefore would be unreasonable to use the same accuracy guidelines for participants at this age as we do with adults. However, by loosening the eligibility restrictions, I am effectively changing the comparison between the two groups. By design, all adult participants accurately identified both French and Mandarin at above chance performance, which allowed us to focus on comparing their performance across the two conditions of interest. For the children, I begin by asking whether they were able to continuously identify the language being spoken, before testing if the condition of a given passage affected their identification ability.

Participants

Results

Descriptive statistics of children's responses show that on the whole they performed at chance accuracy. The table below shows average accuracy, correct rejection, and correct moves scores for each language and condition.

Table 4: Experiment 3: Scores by condition for all three dependent variables

	Accuracy	*Correct Rejections*	*Correct Moves*
French - overall	**0.454 (sd = 0.156)**	**0.560 (sd = 0.166)**	**0.490 (sd = 0.201)**
French – unnatural	0.446 (sd = 0.165)	0.575 (sd = 0.165)	0.478 (sd = 0.218)
French – naturalistic	0.462 (sd = 0.148)	0.544 (sd = 0.167)	0.501 (sd = 0.183)
Mandarin – overall	**0.480 (sd = 0.167)**	**0.539 (sd = 0.159)**	**0.543 (sd = 0.194)**
Mandarin – unnatural	0.486 (sd = 0.168)	0.537 (sd = 0.159)	0.531 (sd = 0.189)
Mandarin – naturalistic	0.473 (sd = 0.168)	0.540 (sd = 0.160)	0.554 (sd = 0.199)
Grand Total	**0.467 (sd = 0.162)**	**0.549 (sd = 0.163)**	**0.516 (sd = 0.199)**

To begin, I assessed the distribution of all three dependent measures. The accuracy and correct rejection measures are approximately normally distributed. The correct moves measure has a skewness of 0.17, but does have relative peaks at 0 and at 1 such that it is not truly normally distributed.

As an initial analysis, I compared all three dependent measures, in each of the four language * condition interactions, to chance performance (0.50) using one-sample two-tailed t-tests, in order to assess whether children's performance differed from chance.

Accuracy:

Participants' accuracy in Mandarin did not statistically significantly differ from chance performance in the unnatural ($t(112) = -1.12$, $p = 0.266$, 95%CI: [0.451, 0.514]) or the naturalistic ($t(111) = -1.149$, $p = 0.14$, 95%CI: [0.445, 0.508]) conditions. Accuracy in French was actually significantly *worse* than chance performance (unnatural: $t(112) = -3.853$, $p < 0.001$, 95%CI: [0.410, 0.471]; naturalistic: $t(111) = -2.423$, $p = 0.017$, 95%CI: [0.439, 0.494). This comparison no longer reached significance in the naturalistic condition after adjusting the alpha-level for multiple comparisons using the Bonferroni method, though the unnatural condition in French remained significantly worse than chance performance even after adjustment.

Correct Rejection:

Participants correctly rejected Mandarin when Mandarin was not playing at levels slightly higher than chance. This comparison did not reach significance in the unnatural condition after adjusting for multiple comparisons ($t(112) = 2.211$, $p = 0.029$, 95%CI: [0.503, 0.563), though it did in the naturalistic condition ($t(111) = 3.05$, $p = 0.003$, 95%CI: [0.516, 0.575]). They also correctly rejected French when French was not playing greater than chance, and both comparisons remained significant after adjusting for multiple comparisons (unnatural: $t(112) = 4.688$, $p < 0.001$, 95%CI: [0.542, 0.604]; naturalistic: $t(111) = 3.10$, $p = 0.002$, 95%CI: [0.518, 0.580]).

Correct moves:

Participants' moves into the Mandarin box were above chance during the naturalistic condition ($t(111) = 3.519$, $p < 0.001$, 95%CI: [0.528, 0.599), but not during the unnatural condition ($t(112) = 1.556$, $p = 0.12$, 95%CI: [0.493, 0.562]). This comparison during the naturalistic condition remained significant after correcting the alpha-level for multiple comparisons among switches into Mandarin. Moves into French were not above chance in either the unnatural ($t(112) = -0.946$, $p = 0.35$, 95%CI: [0.440, 0.521]) or naturalistic conditions ($t(111) = -0.215$, $p = 0.83$, 95%CI: [0.463, 0.530]).

The above results confirm suspicions based on visual inspection of the data: participants' performance hovered around chance, though they did exceed chance accuracy in the correct rejection and correct moves dependent measures in a subset of conditions. Unsurprisingly, performance was best in the correct rejection measure. As previously discussed, this measure may be somewhat inflated due to some participants' tendency to spend large portions of a story with their mouse outside both target boxes, therefore accumulating time toward the correct rejection score for both languages with no apparent effort. In fact, a participant could have received a perfect correct rejection score for a story by doing nothing and simply leaving their cursor outside both boxes for the entire duration of the story (nobody did this, but it was possible).

In order to test the effects of language and condition for each of these three dependent measures, I fit linear mixed effect models. In all models, I additionally accounted for age as a possible interaction with language and condition, given findings in previous research that 7-8 year-olds may be better at discriminating two unfamiliar languages than 4-5 year-olds. All models used a random effects structure with random slopes of language nested within participant. Models failed to converge when story was added as a random intercept.

Accuracy and correct rejections:

The model using accuracy as a dependent variable included no significant or marginally significant fixed effects. This pattern remained true when excluding age from the model and including only language, condition, and the interaction between language and condition. The same was true for models using correct rejection as the dependent variable: no fixed effects were significant, in either the maximal or simplified models. There is no indication that participants' accuracy differed for Mandarin vs. French, or for the naturalistic vs. unnatural conditions.

Correct Moves:

For the model using correct moves as a dependent variable, there was a significant effect of language such that participants' moves into a box were correct more often when that destination box was Mandarin rather than French, $t(36.65) = -2.79$, $\beta = -0.02845$, $p < 0.01$. This effect held when age was removed from the model to focus on language and condition, $t(37.61) = -2.822$, $\beta = -0.029$, $p < 0.01$. However, there was no significant main effect of condition nor an interaction between language and condition. There is therefore no evidence that participants performed any differently when they were hearing naturalistic as opposed to unnatural acoustic cues to code-switching. These results from the correct moves measure should be interpreted with caution due to the distribution of the variable; it is symmetric about the mean and is not a binomial response as results could vary continuously from 0 to 1, but 0 and 1 are local peaks thanks to stories in which participants moved their mouse a very small number of times.

Exploratory analyses: age effects

In the models described above, I found no main effect or interaction with children's age (as a continuous variable). However, as an exploratory analysis inspired by previous findings that 7-8 year-olds may be better at discriminating unknown languages than 4-5 year-olds, I directly compared performance from the oldest and youngest participants. Participants in the experiment 4 ranged in age from 5.08 – 9.08 years of age; I therefore divide them into three approximately equal-sized groups: a young group (5.08-6.41, n = 14), a middle group (6.42-7.38, n = 13) and an older group (7.39-9.08, n = 13). Independent samples t-tests with unequal variance comparing the young group and the older group do not show statistically significant differences in their overall accuracy ($t(285.48) = 1.49$, $p = 0.14$, 95%CI: [-0.009, 0.063]), correct rejection accuracy ($t(287.93) = 0.26$, $p = 0.79$, 95%CI: [-0.033, 0.042]), or proportion of correct moves ($t(297.95) = 0.044$, $p = 0.97$, 95%CI: [-0.043, 0.045]). There is no evidence that the older children were better than the younger children at this task.

<u>Discussion</u>

In Experiment 3, I found that 5- to 8-year-old children overall were unable to reliably and accurately continually identify which of two unfamiliar languages they were hearing. I also found no reliable evidence that this ability differed between the natural and the unnaturalistic conditions; behavior hovered around chance in both conditions, and did not significantly differ between conditions. I will begin by discussing children's difficulties with this particular task, and then the lack of an effect of stimulus condition.

Difficulty performing the task

The finding that children could not reliably discriminate French and Mandarin is perhaps not surprising, especially given previous research on this subject. Although previous studies have found some evidence that children could discriminate unfamiliar languages, whether spoken by a single speaker or two separate speakers, (Bond & Stockmal, 2005; Marks et al., 2009; Potter & Saffran, 2015; Stockmal, 1995; Stockmal et al., 1994), such abilities are not universal. Not all children were able to perform such discrimination tasks in previous research, and younger children (~4-5 years old), in particular, appeared to struggle. Some of these previous patterns of results suggested that younger children relied on an identity-matching heuristic, rather than conceptual representations of each language and comparing the current token to each representation. That is, in an AX same-different task, younger children were better at making "different" judgments than "same" judgments. Further, an unexpectedly long pause separating two otherwise-identical passages led to an increase in "different" judgments for 4-5 year-olds only, compared to when the latency between two identical passages was shorter (Stockmal et al., 1994). While 7-8 year-olds may perform similarly to adults when two languages are produced by different speakers, only some appear to be able to accurately identify whether two passages produced by the same bilingual speaker are from the same or different languages (Bond & Stockmal, 2005; Marks et al., 2009).

While the task in Experiment 3 differs slightly from these AX same-different tasks, as participants were told what the two languages were and that the speaker would switch languages sometimes, it requires a similar or greater degree of meta-linguistic awareness and judgment. To succeed, a listener must have some representation of French and of Mandarin and be able to compare the current auditory input to speech that could be generated by each language. Their task is to identify the language which is most likely to have generated the current signal. Some of the children who participated in this study but performed at chance levels of accuracy may have lacked detailed enough representations of French and Mandarin to continually engage in this comparison. Others may have had the linguistic knowledge but misunderstood or been unable to perform the task. Some only moved their mouse between boxes once or twice during a single passage: perhaps they believed that the language would only switch once or twice in each story, rather than 5-8 times. They may also have had difficulty in manipulating the mouse, which would be reflected in either barely touching it or moving it often without regard for the auditory stimuli. However, as these unusual behaviors with the mouse were more common in Experiment 3 than in the pilot study using English-Mandarin, it is unlikely that difficulty with the mechanics of the task account for the entire pattern of results.

It is also possible that we do not see an overall ability to identify the language due to using only English monolingual children in experiment 3. Perhaps bilingual children (or even monolingual children who live in a highly multilingual community), even if they are not familiar with the specific target languages, would have more meta-linguistic awareness due to their exposure to multiple languages. Bilingual children, especially by age 7-8, may have a deeper understanding of the fact that multiple languages exist and that these languages can differ in their general sound patterns, than children in predominantly monolingual environments. However, Marks et al. (2009) did not find any difference between children who were English monolinguals, L1-Spanish speakers learning English in school, or Spanish-Basque bilinguals learning English in school in the ability to perform same-different judgments on pairs of unfamiliar languages spoken by a single bilingual. Based on that finding, then, it is somewhat unlikely that including a sample of Spanish-English bilingual children in this study would have meaningfully changed the pattern of results. Furthermore, while there was an interaction between code-switching frequency and stimulus condition in the sample of adults in Experiment 1, bilinguals were not overall faster or more accurate than monolinguals. In fact, while not statistically significant, monolinguals may have been more accurate than bilinguals:

7 bilinguals but 6 monolinguals were dropped from analysis due to low accuracy, and an average of 1.05 trials were dropped from analysis from each of the 40 bilinguals with sufficient accuracy while only 0.77 trials were dropped from analysis from each of the 40 monolinguals with sufficient accuracy.

Lack of an effect of stimulus condition

As discussed, I found no evidence of a difference in children's performance in the naturalistic and unnatural conditions. This result is similar to that found for the monolingual adults, though it is important to note that I only included data from adults who were able to identify both languages at above-chance accuracy. However, this pattern suggests that monolingual listeners, who have minimal experience both listening to code-switching and managing phonological representations of two languages, are not sensitive to sub-phonemic distributional shifts in speech sounds. In the discussion section for experiment 1, I considered this lack of sensitivity as a potential disadvantage, in that monolingual listeners may be less aware of the phonological detail of speech in unfamiliar languages and unable to use such detail to anticipate a language switch. However, an alternate framing of this same null effect is that it implies the naturalistic cues to code-switching are not harmful for listeners, even if those cues and the languages from which they are derived are unfamiliar.

One possible consequence laid out in the introduction was that the "blending" of phonetic properties between two languages at the point of a code-switch due to natural acoustic cues to code-switching could make naturalistic code-switches actually *more difficult* to detect for listeners unfamiliar with the two languages. That is, I expect listeners to have some rudimentary representation of both Mandarin and French, due to a combination of prior life experience and the familiarization trials in this experiment in which YM tells a minute-long story in Mandarin and in French while the relevant language is labeled. Though the prior experience differs among individuals, and I expect adults to have stronger baseline familiarity with Mandarin and French than children, the introduction to YM's voice and his variety of each language when produced in a unilingual context is established in a consistent manner for all participants. As this familiarization to YM comes from a unilingual language mode without natural cues to code-switching, and I likewise expect that any life experience participants have with French and Mandarin would not have included code-switching between the two languages, participants should not have heard acoustic cues to French-Mandarin code-switching prior to the first naturalistic trial. Participants' existing representations of French and Mandarin, including the variability in rhythmic properties, intonation, and the acoustic properties of individual speech sounds, should therefore be based on unilingual speech. If these representations are sufficiently detailed, then we might expect that a production which deviates from either language-specific distribution in a way that leads to relative ambiguity about the language from which it was derived would make language identification more difficult.

In the context of the broader goals of this dissertation, such a possibility is particularly relevant in considering the development of these language-specific representations. If code-switching makes it difficult for listeners with highly limited lexical knowledge to track which language is being spoken not simply because two lexical and syntactic systems are present, but because the languages converge slightly at the phonetic level, then the resulting language input environment may be one that provides less clear guidelines both for distinguishing and tracking each language and for the typical properties of each language. Infants and other language learners may have difficulty both identifying which language is being spoken using their existing perceptual discrimination abilities and biases, including discriminating individual phonemes and rhythmic patterns and tracking distributional information about each language in a relatively specified and encapsulated manner.

However, the current results do not support the above interpretation, at least for adults and children. Instead, it seems that listeners who do not know the two languages are often insensitive to these sub-phonemic acoustic cues and that the presence of such cues is neither helpful nor harmful. Perhaps when phonological representations of each language are weak, listeners (particularly those with limited experience with code-switching) do not have detailed expectations regarding the nature of how phonetic variability in each language could co-occur with code-switching.

In experiment 4, I continued to test these same effects in English monolingual infants. Experiment 4 focuses on infants both for practical reasons (the present questions about language learning and development using solely speech input without structured pedagogy or a focus on lexical items are most akin to an infants' environment and developmental needs) and for theoretical reasons (infants show a robust ability to discriminate language passages on the basis of rhythmic information, and to discriminate a wide range of speech sounds prior to 10 months of age). Though the monolingual infants are even less likely than monolingual children to have any prior exposure to French and/or Mandarin, their perceptual skills are broadly applied to all speech input without an expectation of detailed understanding. It is therefore possible that infants might both show a better baseline performance than children, and show more sensitivity to the naturalistic – unnatural comparison due to this perceptual awareness. Infants' performance may have also been boosted relative to children's due to the differences in task, as infants first participated in a training phase which provided non-contingent feedback on a given speech sample's language membership. Further, the task used with infants does not require the same degree of explicit meta-awareness regarding language *identification* as the task used in Experiments 1-3.

The details of the method, including the design of this training phase, will be described further below. For now, the crucial point to note is that while the children are told "this is Mandarin" while hearing a minute-long story in Mandarin (and same for French), the infants, over the course of the experiment, are repeatedly trained in the difference between Mandarin and French. When Mandarin is playing, they see images on one side of the screen (or no images at all). Those images appear on the other side of the screen when French is playing. The infants therefore receive continual training, both prior to and during the target trials, while the children and adults received a single minute-long familiarization trial for each language and no subsequent feedback. Direct comparison of performance between the children and the infants is therefore impossible. I will return to this difference after describing the methods and results for experiment 4.

Experiment 4: Infants

Introduction

In Experiment 4 I tested infants' ability to distinguish between two unfamiliar languages in the presence of code-switching, and crucially whether the presence of naturalistic cues to code-switching impacts their ability to make this distinction.

Recall that as previously discussed, both infants and adults show a robust ability to use rhythmic and durational cues to discriminate between languages, including pairs of unfamiliar languages. Though the minimally-required distance in rhythmic properties for infants to discriminate languages has not been formally established, monolingual and bilingual infants exposed to Spanish and/or Catalan are able to discriminate between these two languages which are both traditionally categorized as "syllable-timed" (Bosch & Sebastián-Gallés, 1997; Bosch & Sebastián-Gallés, 2001). Spanish and Catalan monolingual infants may also be able to detect that a switch between these two languages has occurred (Birulés et al., 2024). English monolingual infants can also discriminate between some pairs of rhythmically similar languages (Spanish vs. French) and varieties of their native language (Plymouth English vs. Welsh English) but not all (Spanish vs. Finnish; French vs. Finnish; Welsh English vs. Scottish English; White et al., 2014, 2016). It therefore appears that infants are sensitive to aspects of durational variability between samples of languages. Adults can make similar fine-grained distinctions, discriminating between varieties of English which have been resynthesized to remove deterministic segmental information (White et al., 2012).

Importantly, these findings with both infants and adults have relied on methods that require some combination of familiarization and off-line categorization judgment. These methods (with the partial exception of Birulés et al., 2024; discussed further below) do not reflect the real experience of listening to a person who speaks multiple languages and needing to identify the language that is being spoken, let alone when that change occurs. Infant studies use several methodologies, with choice depending partially on the participants' age. The most common paradigm for newborns, the high-amplitude sucking paradigm (HASP), involves measuring infants' habituation to a particular language using their instinct to suck on a pacifier placed in their mouth. When the habituation criteria is reached, the language is changed for the experimental group only and pacifier sucking amplitude is compared for the pre- and post-stimulus change time windows. For older infants with sufficient head and neck control a common methodology in this field is the familiarization-preference procedure where infants are familiarized with a language and then their preference for listening to that familiarized language is compared to that of a different, unfamiliar language through a series of trials.

HASP and similar methods involving habituation typically include a maximum of one language change per participant and test whether infants' listening behavior differs before and after that language change; if it does not change, we assume that the acoustic difference between the languages was not sufficiently noticeable or interesting to the infants to regain their attention. For methods based on familiarization-preference, the language may change between trials, which are offset with an attention getter involving flashing lights and/or sounds. However, experimenters do not measure this trial-to-trial difference in infant attention, but rather their overall looking time to each language regardless of the immediately preceding trial type. Thus, these methods test for *acknowledgement of a difference* but not *the moment when a change in some continually appearing stimulus is detected.*

Two recent studies that aimed to more explicitly test infant's recognition of language *switching* rather than *discrimination* between languages come closer to answering this question, but without fully addressing the question of the amount of time that infants need in order to detect a language switch. Schott, et al. (2021) used the basic headturn preference procedure (HPP) to test if infants could distinguish between stimuli presented in a single language vs. presented in multiple languages. Monolingual and bilingual 8- to 12-month-old infants listened to speech presented in English and French (for babies living in Montreal) or English and Spanish (for babies living in New Jersey). When the stimuli were composed of word lists in either a single language vs. switching back and forth between two languages (i.e. "Doggy, Balloon, Doggy, Balloon…" vs. "Doggy, *Globo*, Doggy, *Globo…*"), there was no evidence for different looking times between the two conditions. Further, using sentences which end in a single-word code-switch (i.e. "Do you like the *gato*? I want the *leche*!" for English-Spanish switches) as compared to unilingual sentences (i.e. "Do you like the foot? I want the cookie!") yielded inconclusive results. One interpretation of this finding is that infants in the second half of their first year of life in fact can not distinguish between unilingual and mixed-language speech, regardless of whether they regularly hear multiple languages in their environment.

An alternative interpretation is that the methodology used in Schott, et al. (2021) was not able to test infants' capabilities to solve this particular problem. The authors note that an equivalence test comparing their observed effect size to a minimal detectable effect size was not statistically significant, which could indicate the need for a larger sample size. However, aspects of the stimuli and test method themselves could have also come into play. This study, unlike those previously discussed, did not aim to test whether infants could distinguish the language in a passage from one that had previously been familiarized. Instead, it took as inspiration the finding that infants tend to be more interested in two sounds that alternate as compared to one that repeats, and aimed to test whether this pattern would extend to words in two languages. It is possible that infants noticed that the unilingual and language-alternating trials were different but that the samples were equally interesting as both contained some degree of variation, leading to comparable looking times between the two conditions.

Alternatively, the infants could have failed to distinguish between unilingual and mixed-language stimuli particularly because of the use of word lists and sentence-final single-word insertions, as compared to more naturalistic situations. As these types of switches involve short durations of speech in at least one of the languages, they may effectively remove the possibility of using rhythm to distinguish them. Although durational characteristics such as %V can be measured on the basis of a single word, such measurements are likely to be more accurate and robust to variation that could arise from single language-atypical words when they are instead computed over longer durations. In the Schott, et al. (2021) paradigm, infants instead needed either sufficient phonological distinction between the languages or knowledge of which language category a given word belongs to. While some evidence suggests that older toddlers (starting at age 18 months) do detect a sentence-final switch consisting of a single-word (Byers-Heinlein et al., 2017; Morini & Newman, 2019; Potter et al., 2019) and therefore have some of this requisite phonological and lexical knowledge, the vocabularies of 8- to 12-month-old infants are relatively limited.

Using a different approach, the habituation-switch method, Birulés et al., (2024) tested Catalan-Spanish monolingual and bilingual 4-month-olds' ability to identify a switch between Catalan, Spanish, and English using audiovisual information. Infants were habituated to a video of a person speaking their dominant language (either Catalan or Spanish) and then saw a series of three trials, each consisting of two 10-second passages with no intervening break or attention getter. In the first trial infants heard and saw the same language to which they had just been familiarized, in the second they heard and saw the other language (either Spanish or Catalan; "close switch"), and in the third trial they heard and saw English ("distant switch"). Infants who were exposed to only Spanish or Catalan at home attended to both the close switch trial and the distant switch trial longer than same language trial, indicating that they did notice a switch between Spanish and Catalan. Though trends indicated that they may have looked longer to English than the close switch trial, that result wasn't significant. Further, noticing the switch between Spanish and Catalan appeared to take infants more than 10 seconds, as their looking time only exceeded that of the same language trial in the second 10-second passage of the close switch trial, but not the first. Infants who heard both Spanish and Catalan at home looked longer to English than to the same language trial or the close switch trial (especially in the second 10-second passage). However, they did not attend to the screen more following a switch between Spanish and Catalan than to the same-language trial, and therefore may not have noticed this switch between their two familiar but rhythmically similar languages.

This study leads to several conclusions about infants' abilities to distinguish between languages in an ongoing passage which are relevant to the current research question. First, and most importantly, the authors found evidence that infants can respond to multiple sequential language switches without an intervening break. The monolingual infants, following habituation and the same-language trial, attended to the screen for a longer amount of time in the close switch trial and then for an even longer amount of time in the distant switch trial. Although other habituation-test paradigms have also shown that infants can respond to a change between two types of passages, such paradigms are primarily used with newborns and use high-amplitude sucking, rather than with somewhat older infants in a looking-time procedure, and are rarely successful at testing more than one switch.

However, several questions remain based on this finding. First, although this paradigm is a more accurate assessment of language switch detection than familiarization-preference procedures (which have previously shown infants' ability to discriminate rhythmically similar languages such as Catalan-Spanish or Spanish-French), it is still a test of whether infants notice that some change *has occurred* rather than *when that change occurs*. That is, by measuring their looking time during the entirety of the close-language and distant-language switch trials, and then comparing that looking time to that in the same language trial, the authors show that these monolingual 4-month-olds can tell that a 10-second passage of Catalan differs in some meaningful way from a 10-second passage of Spanish. Although there is some indication that infants need 10 seconds of exposure to the post-switch language to notice a change between similar languages, the experimental design does not allow for a detailed measure of reaction time. We do not know, based on these results, how long it took infants to notice a language switch.

Further, there are several possibilities regarding the cues that infants could have used to notice a language switch in this paradigm. The use of solely audiovisual stimuli rather than audio-only makes it difficult to determine whether infants relied solely on acoustic differences between the languages (including differences in rhythmic or durational properties, as well as segmental information such as phonemic inventory and phonotactic regularities), or if the addition of concurrent mouth movements was necessary to support discrimination. The finding that monolinguals, but not bilinguals, dishabituated to a switch between Spanish and Catalan while both groups looked longer to English than to their dominant language also raises questions about the specificity of infants' representations of the language(s) they hear most regularly. That is, perhaps these four-month-old Catalan-Spanish bilingual infants, despite being able to distinguish between longer passages of the two languages, struggle to notice a language switch when the languages are produced in close temporal proximity, and therefore do not fully separate their language input into different "streams" while listening. Monolingual infants, on the other hand, hear just one of these two languages regularly and may therefore have a stronger, more detailed, representation of the rhythmic and segmental properties of that one language. This more specified representation could support their ability to detect when a switch occurs, even if that switch results in a small degree of change between those properties. Both groups of infants may essentially be responding to a change that results in unfamiliar input. If that is the case, then we would expect results to be similar in a version of this paradigm that used French instead of English as the third, unfamiliar language, with both monolingual and bilingual infants looking longer to this unfamiliar language which has similar rhythmic and durational properties to both Spanish and Catalan but has some differences. While the current study does not aim to directly answer these remaining questions or assess how bilingual infants detect a language switch between their two languages, this discrepancy between monolingual and bilingual Catalan-Spanish infants is important and will be returned to in the discussion.

The current study offers a different approach to addressing the detection of language switching among young infants. I used an adaptation of the anticipatory eye movement paradigm (McMurray & Aslin, 2004) to train infants to associate one direction with Language A and the other direction with Language B. I then tested the extent to which they look to the correct direction following a language switch. The anticipatory eye movement (AntEM) paradigm has previously been used to test shapes, words, syllables (/ma/-/na/) and phonemes (/e/-/u/, /ɛ/-/e/) (Albareda-Castellot et al., 2011; Kay & Adler, 2023; McMurray & Aslin, 2004). A similar paradigm was also used in Kovács & Mehler (2009b), which exposed infants to two linguistic structures (ABA and AAB) and trained them to associate each structure with a particular direction. A clear benefit of this paradigm over many others is that it includes distinct responses for two types of stimuli and can therefore test categorization as opposed to simply discrimination. As discussed by Albareda-Castellot and colleagues (2011), such a dual-response paradigm can be more sensitive than one that relies on listening time as a measure of preference. However, these alternative forced-choice paradigms also often have lower success rates than those driven by listening preference, as infants must be trained to perform distinct reactions in response to different types of stimuli; this training requires a relatively higher level of attentional control and cognitive functioning.

Stimuli

In Experiment 4, infants were trained to produce different behaviors in response to French and Mandarin Chinese. Following McMurray & Aslin (2004), the experiment is split into a training phase and a test phase. In the training phase infants were taught to associate each language with looking in a different direction. Mapping of language to direction was counter-balanced across participants, though I will use the French-left Mandarin-right mapping here for the sake of clarity.

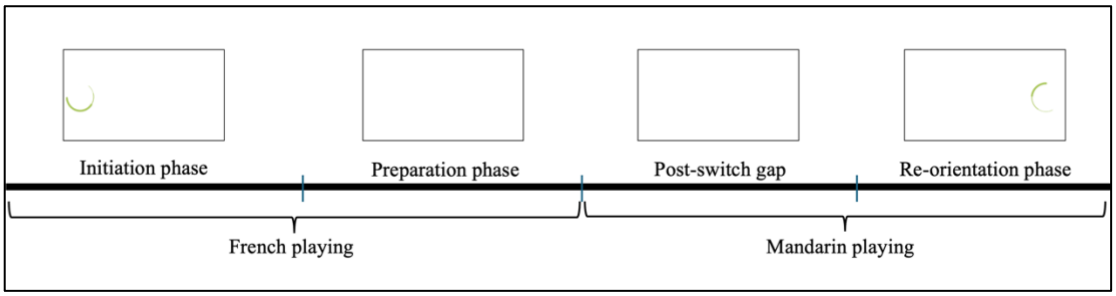
This training was accomplished with the use of moving images that appear and disappear on either the left or the right side of the screen: when French is playing, the image is either on the left or not present; when Mandarin is playing, the image is either on the right or not present. Over the course of the training phase the delay before the moving image reappears on the screen gradually increases, to condition infants to look to the appropriate side in anticipation of the image appearing. Once the delay has grown to 1840 ms, the experiment then automatically moved into the test phase. The test phase consisted of 24 trials with a consistent 2000 ms delay. Infants' anticipation of the moving image was tested by measuring the degree to which they preferentially looked in the direction associated with the language that was currently playing. The following section begins with a more detailed description of the stimuli, followed by an explanation of the participants and data collection and analysis procedures.

The task consists of a total of 48 trials: 24 in the training phase and 24 in the test phase. I use both "stay" trials, in which the language does not switch, and "switch" trials, which switch between Mandarin and French at the approximate mid-point. The use of stay trials ensures that infants do not learn a pattern that the moving image will always switch sides in the middle of a trial. Instead, in order to correctly anticipate the location of the moving image, they need to discriminate the languages and identify which is currently playing. There were thus four trial types: French-only, Mandarin-only, French-Mandarin, and Mandarin-French. All trials were 5-8 seconds long and had the same four-part structure. A layout of this structure is also shown in Figure 12 (using a French-Mandarin switch trial as an example).

1) Initiation: For the first 2000ms, the moving image was visible in the appropriate side of the screen (left for French-only and French-Mandarin, right for Mandarin-only and Mandarin-French)
2) Preparation: 2000ms into the trial, the moving image disappeared, but the sound continued to play in the same language as before. The preparation phase lasted for a varying duration, until the language switch (for switch trials) or a matched mid-point (for stay trials). The screen was blank during the preparation phase. The preparation phase was designed so that infants could not predict when the language switch would occur based on the disappearance of the moving image.
3) Post-switch gap: The sound either switched languages (for switch trials) or continued in a single language (for stay trials); this point marks the onset of the period referred to as the post-switch gap. The screen was still blank during the post-switch gap. The duration of this segment is determined by the trial number and phase. During the training phase, this gap increased from 0 to 1840ms, by 80ms each time. Thus, the gap was 0ms during training trial #1, 720ms during trial #10, 1520ms during trial #20, and 1840ms during trial #24. The gap lasted for a consistent 2000ms during all 24 trials in the test phase.

4) Re-orientation: After the post-switch gap, the moving image reappeared and remained visible for the remainder of the trial. The image switched sides from its original location for the switch trials, and re-appeared in the original location for the stay trials.

Figure 12: Experiment 4: Outline of the structure of each trial (switch trial). Single frames from the initiation and re-orientation phases are shown here, but during the experiment the image was moving (i.e. rotating, growing and shrinking in size, flashing)



The physical characteristics (shape, color) of the moving image changed for every trial, so that infants could not habituate to the image itself and become bored. Moving shapes were obtained from Rocketstock.com and their color, size, speed, and location were manually adjusted in Adobe Premiere Pro. All shapes were set to the same size and to appear at the far-left or far-right of the screen.

Trials were offset using either a 500ms black screen or a longer attention getter video. Attention getter videos appeared every four trials during the training phase. They appeared every two trials or every four trials during the test phase, as I discovered part-way through data collection that some infants had trouble attending to the screen for four consecutive trials by the end of the experiment. Additional attention getter videos were therefore added to the test phase for later participants in an attempt to retain their engagement in the study for a longer period of time. This attempt was not always successful.

Audio stimuli were created by selecting short segments from the stimuli for experiments 1 and 3. Half of the infants were assigned to the naturalistic condition, in which the audio stimuli for both the training and test phases were derived from the naturalistic passages described in the overall methods section for this set of experiments. The other half of infants were assigned to the unnatural condition and heard clips of stories which were equivalent in content, but came from the unnatural passages and thus differed acoustically. Condition (unnatural vs. naturalistic) was therefore a between-subjects variable in experiment 4, while it was a within-subjects variable in experiments 1 and 3. All infants heard the clips in an identical, fixed, pseudorandomized order. The only restriction placed on this order was that the same language would not appear in three consecutive segments. That is, a Mandarin-only trial could be preceded by a French-Mandarin switch trial or followed by a Mandarin-French switch trial, but not both, such that the child could hear either … FM MM … or …MM MF …, but not … FM MM MF … . Similarly, two stay trials of the same language were not presented consecutively (… FF FF …; … MM MM …).

As the stimuli in the naturalistic and unnatural conditions cover identical content but stimulus durations were unequal due to minor differences in speech rate between the two conditions, the time points at which the switch points (for switch trials) or midpoints (for stay trials) occur differ slightly. Table 5 below shows the average latencies at which the switch point (or matched midpoint) occurred in both the training and test phases for the unnatural and naturalistic conditions. Two-tailed independent samples t-tests with unequal variances indicate that the latency of this switch point between the two trial types did not differ in any of the conditions or phases of the experiment. Appendix VII includes a description of all stimuli used in this experiment including trial type, the passage that the audio file is taken from, the latency of the switch point, and the length of the post-switch gap.

Table 5: Experiment 4: Average latency of switch points for all trial types

	Switch trials	***Stay trials***	
Training phase			
Unnatural condition	3.45 seconds	3.47 seconds	t(19) = 0.12, p = 0.91
Naturalistic condition	3.50 seconds	3.61 seconds	t(21) = 0.46, p = 0.65
Test phase			
Unnatural condition	3.66 seconds	3.71 seconds	t(21) = 0.24, p = 0.82
Naturalistic condition	3.72 seconds	3.88 seconds	t(15) = 0.73, p = 0.48

Procedure

Infants were recruited using Lookit/Children Helping Science (Scott & Schulz, 2017) and by contacting families through the University of Maryland Infant and Child Studies database. Interested families were directed to select a time to participate using Calendly. Families could be located anywhere in the United States; they were not required to be local to the University of Maryland. All participating families received a $5 online gift card via Tango as compensation. At the time of the appointment, families joined a Zoom meeting with an experimenter (the author) using a laptop or desktop computer. The experiment was run via a PowerPoint presentation which the experimenter screen-shared using Zoom. The experimenter discussed the consent form with the family and obtained their consent by asking the parent to read a statement of verbal consent while the experimenter recorded the Zoom meeting. The experimenter then engaged in a series of set-up checks with the family to ensure that the study would run as designed.

First, the experimenter ensured that the family was unlikely to be interrupted in the next 10-15 minutes. This check was only moderately ineffective, as several families were interrupted by pets, older children, or adults in the home, though most did attempt to complete the study in a quiet location. The experimenter then asked the participant to fully turn up the screen brightness on their device and showed all-white, all-black, and all-red screens to check that the difference between these three colors was visible when reflected on the participant's face. If color changes were not visible, the family was asked to adjust lighting sources in their home as much as possible and the color check was re-tested. In a small number of cases, lighting could not be adjusted to allow the experimenter to assess which color was reflecting on the participant's face; data collection was continued anyway for the sake of participant experience, though most of these cases were unable to be properly scored. The experimenter also ensured that the participant's screen was set up appropriately. Participants were instructed to put Zoom in full screen mode and standard view, so that the presentation that was being screen-shared took up the entirety of their screen and both the participant's and experimenter's video feeds appeared in a movable window. They then minimized this window and moved it to the side of the screen so that it would not be visible during the study. The parent then either sat with their infant on their lap, in which case they were told to keep their eyes closed as much as possible, or put their child in a highchair and stood next to or behind their child so that they were outside of the child's direct line of sight. The experimenter's microphone was muted to avoid interrupting the study, though occasionally the experimenter needed to unmute in order to communicate with the parent or say the child's name to regain their attention. The experimenter began recording the Zoom session once the participant and experimenter were ready.

The experiment began with a brief calibration phase, during which silent large moving images appeared on the left then the right sides of the screen; each moving image continued to play until the experimenter judged that the infant was attending to the screen and looking at that image. This phase was not scored, but could be used by the coders to establish the boundaries of a left and right look for that particular child. As families naturally varied with respect to the physical distance between their child and the size of the computer screen on which they viewed the study, the visual angle between the two sides was not equivalent across children. The use of this calibration phase allowed us to assess that the child will look to both directions, and to approximate the visual angle of each of those directions for them, independently of the experiment itself.

Data Scoring and Analysis

All data was coded frame-by-frame by highly trained undergraduate research assistants using Datavyu. Scorers used the lighting color changes to detect when each trial began and ended, as all trials used a white background and all attention getter or inter-trial breaks used a black or red background. For the first 16 trials, scorers did not code the direction that the child was looking, as these trials were used to teach the child the language-direction mapping and provide them with some degree of training regarding the difference between French and Mandarin, but not to test the degree of their anticipatory eye movements. Scorers then coded whether the child was looking to the left, right, or off-screen during the remaining 32 trials (the last 8 training trials and all 24 test trials). Each video was only coded by one scorer during this initial analysis, though 25% of videos will also be coded by a second scorer to establish reliability.

Following scoring, the coded start and stop times of each trial were checked against the "ground truth" length of each trial based on the video file associated with the trial. If the length of a coded trial (based on color changes) differed from the expected length of that trial by more than 100ms, the author then looked at the video to establish reliability for the trial start and stop times. In most cases, this second coding was sufficient to adjust the length of the trial to be within 100ms from the expected length. However, in some cases the author agreed with the first coder and the coded trial length continued to differ from the expected length. This analysis excludes trials whose coded length differs from their expected length by more than 200ms after examination by the author. A discussion of possible explanations for this occasional discrepancy is included in the limitations section. Assessing coded trial length also allowed for a check of undiscovered experimenter error. For one participant the experimenter was found to have mistakenly shown a pair of trials in the test phase twice. The second presentation of these trials was excluded from analysis. For another participant, four of the first sixteen trials during the training phase were mistakenly skipped. This participant is discussed in the results section.

After establishing that the number and length of coded trials was accurate, I then analyzed the participant's looking behavior during the time region of interest. This region coincides with, but is not equivalent to, the post-switch gap. It begins 350ms after the onset of the post-switch gap, to allow participants time to hear the speech and to initiate and execute an eye movement, and ends 150ms after the offset of the post-switch gap. Although infants' saccade latency in response to a visual stimulus likely exceeds 150ms (Luna et al., 2008), this conservative period of analysis was selected to ensure that analyzed looking time is in anticipation of, rather than response to, the reappearance of a visual stimulus. The analysis period thus had a duration of 1800ms during the test phase and ranged from 1080-1640ms in the eight analyzed trials of the training phase.

I then evaluate the child's looking behavior during this analysis period for each trial. I begin by establishing that the child did attend to the screen during the analysis period; trials with no looks during the analysis period are disregarded. In most cases this occurred because the child was fully disengaged or unhappy during the trial and was also not looking at the screen before or after the analysis period, though occasionally the child was otherwise engaged but happened to look offscreen for the entirety of the analysis period. I then tested the proportion of the child's looking time that was to the left of the screen and to the right of the screen. Previous studies using the anticipatory eye movement paradigm (McMurray & Aslin, 2004; Albareda-Castellot, et al., 2011) scored looking behavior as a binary response; a trial was coded as one with a "left look" if the child looked to the left more than to the right during that trial, and vice versa for a right look. I use a slightly more sensitive measure by calculating the proportion of look duration to each direction, as a function of the total time the child spent looking to the screen. To make it more concrete, take an example trial in which a child looked at the screen for a total of 750ms out of the 1800ms during the analysis period: 525ms to the left and 225ms to the right. Under the analysis method established by McMurray & Aslin

(2004), this trial would be considered a left look, with no consideration for the time the child spent looking to the right. Under the present analysis method, I first established that the child's attention during this trial is 750ms/1800ms = .417. The child's left looking proportion is 525ms/750ms = 0.7 and right looking proportion is 225ms/750ms = 0.3[9].

Results

44 participants between the ages of 5 and 15 months participated in this study (20 male, 24 female, M_{age} = 9.14 months). To date, the data have been scored and analyzed for 17 participants (7 male, 10 female, M_{age} = 9.65 months). 14 additional participants were dropped from further analysis because they were very fussy or had poor lighting/physical conditions that made scoring incredibly difficult; data was partially scored for some of these participants. Data scoring and analysis is ongoing for the remaining 13 participants. Further recruitment and data collection is ongoing to replace the 14 participants whose data was unable to be fully scored, in addition to any participants whose data may be found to be unusable but has not yet been scored. The table below shows demographic information for all 44 participants, as well as their outcome (scoring in progress, dropped, analyzed).

Table 6: Experiment 4: Participant demographics and analysis status.

Status is color-coded for ease of readability: participants with analyzed videos are in green, coding in progress is in yellow, and unanalyzable videos in red

Participant ID number[10]	*Age (months)*	*Gender*	*English % exposure*	*Status*
1	10	F	100	*Analyzed but poor lighting cues*
2	10	F	100	Coding in progress
3	13	M	100	*Poor lighting cues*
4	9	M	100	Analyzed
5	15	M	100	*Analyzed but poor lighting cues*
6	9	M	100	*Poor lighting cues/video conditions*

[9] Using the McMurray & Aslin (2004) analysis method does not change the overall pattern of results. See the results section for an alternate breakdown of participant accuracy.
[10] Several participant ID numbers are skipped because the parent provided informed consent and was thus assigned an ID number, but then decided not to participate at that time because their child was asleep or fussy; attempts to reschedule for these participants have been unsuccessful. A few ID numbers were also mistakenly skipped by the experimenter.

7	11	M	100	Analyzed
8	7	F	100	Analyzed
9	11	F	100	*Poor lighting cues/video conditions*
10	12	F	80-100	Analyzed
11	5	M	100	*Analyzed but poor lighting cues*
12	13	F	100	*Analyzed but child stopped participating early and was very fussy; poor lighting cues*
13	7	M	100	Analyzed
14	10	F	100	*Tough to code; poor lighting cues*
15	14	M	100	*Analyzed but poor lighting cues; baby and mom move a lot*
16	6	M	100	Coding in progress
17	12	F	100	*Analyzed but poor lighting cues and child is fussy*
18	13	F	80-100	Analyzed
19	10	F	100	Coding in progress
20	11	F	100	Analyzed
21	5	M	100	Coding in progress
22	9	M	100	Analyzed
23	8	F	100	Coding in progress
24	5	M	100	*Poor video quality*
26	9	M	100	Coding in progress
27	8	F	100	Analyzed
28	14	F	80-100	Analyzed
29	6	F	100	Coding in progress
31	9	M	100	Analyzed
32	12	F	80-100	Analyzed
33	6	F	100	Analyzed
34	6	M	100	Analyzed
35	7	M	80-100	*Fussy child; occasional poor lighting cues*
36	12	F	100	*Poor lighting cues; baby moves a lot and often hides face*
37	11	M	100	*Poor lighting cues*
38	11	F	100	Analyzed
41	12	F	80-100	Analyzed
42	7	M	100	Analyzed
43	5	M	100	Coding in progress
45	12	M	100	Coding in progress
46	5	F	80-100	Coding in progress
47	5	F	100	Coding in progress
48	5	F	100	Coding in progress
49	5	F	100	Coding in progress

The following section describes the results from the 17 participants whose data has been scored thus far and who did not fuss out of the experiment. The original plan of analysis for this experiment was to run through a series of data reliability checks, before comparing test phase accuracy in different trial types for participants assigned to the unnatural and naturalistic conditions. These planned data reliability checks included tests of attention (did infants attend to the screen for at least 33.3% of the time during the end of the training phase and entirety of the test phase), as well as a test of training success (did infants look to the correct side more than 50% of the time during the end of the training phase). Only infants who showed attention > 33.3% of the time during both experimental phases and accuracy > 50% during the training phase would be included in this analysis of trial types in the test phase. Specifically, I planned to test if there was a difference between infants assigned to the unnatural condition, in which all trials included audio which was recorded unilingually and spliced together, and infants assigned to the naturalistic condition, in which all trials included audio which was recorded with natural code-switches but spliced from separate recordings, in their looking behavior for both the switch trials and the stay trials.

Unfortunately, this series of data reliability checks left us with six participants who attended to the task for at least 33.3% of the time in both the training and test phases, four of whom reached over 50% accuracy in the training phase. The series of tables below shows the breakdown of how each participant fit (or failed to fit) these guidelines. As can be seen in Table 7, 8 of these 17 participants did not attend to the task for at least one third of the analysis period in both the training phase and the test phase; we disregard these participants as their accuracy may be unreliable with such a small period of time used for analysis. Of the 9 remaining participants, 2 attended to the training phase for at least one third of the time but not the test phase, 1 attended to the test phase for at least one third of the time but not the training phase, and 6 attended to both phases for at least one third of the time. In Table 8 also includes information regarding these 9 participants' accuracy in the training and test phases. Table 9 further categorizes these 9 participants on the basis of their attention (greater than or less than 33.3%) and accuracy (greater than or less than 50%) in both the training phase and the test phase. Four participants out of these 17 fully meet the planned criteria for inclusion (and 5 more meet at least some of the criteria).

Table 7: Experiment 4: Proportion of time all 19 fully-analyzed infants attended to the screen during the training and test phases. Attendance > 33.3% in the training phase is highlighted in yellow; attendance > 33.3% in the test phase is highlighted in green.

Participant	***Age (months)***	***Condition***	***Training: proportion of attention***	***Test: proportion of attention***
4	9	Naturalistic	26.4%	9.4%
7	11	Naturalistic	3.1%	18.8%
8	7	Unnatural	49.1%	34.7%
10	12	Unnatural	78.3%	63.6%
13	7	Unnatural	52.3%	46.8%
18	13	Unnatural	12.5%	12.6%
20	11	Unnatural	40.2%	64.3%
22	9	Unnatural	50.0%	18.6%
27	8	Unnatural	75.3%	53.1%
28	14	Unnatural	38.0%	29.1%
31	9	Unnatural	13.4%	12.1%
32	12	Naturalistic	26.1%	5.3%
33	6	Naturalistic	46.3%	33.9%
34	6	Naturalistic	28.7%	12.6%
38	11	Naturalistic	0%	24.4%
41	12	Naturalistic	6.8%	17.4%
42	7	Naturalistic	24.8%	36.5%

Table 8: Experiment 4: Accuracy and proportion of time attending to the screen for the 9 participants who attended to the screen > 33.3% of the time during the training phase or test phase. Attending > 33.3% in the training phase is highlighted in yellow; attending > 33.3% in the test phase is highlighted in green

Participant	***Age (months)***	***Condition***	***Training: prop. of attention***	***Training: overall accuracy***	***Test: prop. of attention***	***Test: overall accuracy***
8	7	Unnatural	49.1%	39.3%	34.7%	69.4%
10	12	Unnatural	78.3%	75%	63.6%	50.7%
13	7	Unnatural	52.3%	61.8%	46.8%	65.5%
20	11	Unnatural	40.2%	81.6%	64.3%	49.5%
22	9	Unnatural	50.0%	64.3%	18.6%	58.1%
27	8	Unnatural	75.3%	58.2%	53.1%	58.3%
28	14	Unnatural	38.0%	46.7%	29.1%	39.5%
33	6	Naturalistic	46.3%	39.3%	33.9%	52.6%
42	7	Naturalistic	24.8%	100%	36.5%	68.4%

*Table 9: Experiment 4: Categorization of accuracy and proportion of time attending to the screen for the 9 participants who attended to the screen > 33.3% of time during the training phase or test phase. Attending > 33.3% in the training phase is highlighted in yellow; attending > 33.3% in the test phase is highlighted in green. *indicates participants whose accuracy is between 49% and 51%*

	***Test:* attending > 33% accuracy > 50%**	***Test:* attending > 33% accuracy < 50%**	***Test:* attending < 33% accuracy > 50%**	***Test:* attending < 33% accuracy < 50%**
***Training:* attending > 33% accuracy > 50%**	3 participants (*10*, *13**, *27*)	1 participant (*20**)	1 participant (*22*)	
***Training:* attending > 33% accuracy < 50%**	2 participants (*8*, *33*)			1 participant (*28*)
***Training:* attending < 33% accuracy > 50%**	1 participant (*42*)			
***Training:* attending < 33% accuracy < 50%**				

On average, participants who attended to the training phase at least 33.3% of the time (all eight of those highlighted in yellow) looked in the correct direction during the training phase 58.3% of the time. They also attended to the test phase 43.0% of the time and looked in the correct direction during the test phase 55.5% of the time[11]. Average accuracy in the switch trials was 61.55% during the training phase and 52.63% during the test phase for these same eight participants. For the six participants who attended to *both* phases for more than 33% of

[11] Using the McMurray & Aslin (2004) method of analysis does result in a greater estimate of accuracy than the present method of analysis, but not a difference in the overall pattern of results. For example, while the current analysis suggests that the eight infants who attended to the screen for at least 1/3 of the analysis period during the training phase had an average looking time accuracy of 58.3% in that training phase, the McMurray & Aslin method of categorizing each trial as correct or incorrect results in of 66.1% of trials being correct in the training phase. In the test phases, accuracy measures for the same group of participants are 55.5% and 56.3%, respectively.

the analysis window, average training accuracy was 59.2% (71.49% in the switch trials only) and test accuracy was 57.67% (58.51% in the switch trials only).

Three participants (*10, 13, 27*) who attended to both phases for at least 33.3% of the time also looked in the correct direction at least 50% of the time during the training and test phases, indicating that they were able to discriminate between French and Mandarin and both learned and maintained the language-side contingency. One additional participant (*20*) attended to both phases for at least 33.3% of the time and looked in the correct direction at least 50% of the time during the training phase, but did not achieve 50% accuracy in the test phase. Participant *20* thus met the pre-set inclusion criteria and initially learned the contingency present in the experiment, but did not sufficiently maintain the ability to exhibit discrimination between French and Mandarin. It is also worthwhile to highlight two participants who attended to both phases for the requisite amount of time and achieved at least 50% accuracy in the test phase but not the training phase (*8* and *33*) and one participant who achieved at least 50% accuracy in both phases and attended to the test phase but not the training phase for at least 33.3% of the time (*42*); although these three participants (*8, 33, 42*) did not reach the pre-set inclusion criteria, they are worth considering given the small sample of participants whose data can be meaningfully analyzed. In the following section I will discuss these seven participants in further detail.

Participants *13* and *27* are 7 and 8 months old, respectively, and both looked in the correct direction at least 58% of the time in both phases. They therefore show the most robust evidence of an ability to learn the relationship between language and direction and to move their eyes in anticipation of where a moving image will appear, including after a code-switch. Their respective accuracy on switch trials during the test phase is 55% and 66.7%.

However, participant *10* (12 months old) looked to the correct direction only 50.7% of the time during the test phase. Although this accuracy value exceeds 50% and participant *10* is thus included in the "accurate" column with participants *13* and *27*, this value is not meaningfully different from chance, or from participant *20* (11 months old) who looked to the correct direction 49.5% of the time during the test phase. For switch trials in the test phase, participants 10 and 20 looked to the correct side 42.9% and 52.4% of the time, respectively. Their accuracy to switch trials in the training phase was 75% and 85.6%. Thus, they both may have learned the relationship early in the experiment but been unable to maintain the ability to continue detecting a code-switch and anticipating the location of a re-appearing moving image. Thus, the two participants who showed the most success during the study were 7 and 8 months old, and the two who showed some early success but did not maintain it were 11 and 12 months old.

Participant *8* (7 months old), who reached accurate looking in the test phase only (69.4% in the test phase vs 39.3% in the training phase) despite attending to both phases over one-third of the time, is a potentially interesting case due to experimenter error during the test phase: four of the first 16 trials in the training phase were mistakenly skipped. This meant that by the eight analyzed trials at the end of the training phase, participant *8* had only been exposed to 12 trials instead of 16. This decreased exposure may have led to her lower accuracy during the training phase, as she had fewer opportunities to learn the contingency in this experiment. However, it appears that later exposure during the test phase helped her to solidify learning of the contingency, as she accurately looked to the correct side during the test phase. Participant *8* also showed a large discrepancy in her accuracy between the switch and stay trials in the training phase, which may further reflect the time needed to learn the structure of the experiment: her accuracy on switch trials was 100% in the training phase (vs 19.06% in stay trials) and was 78.29% in the test phase (vs 4.30% in stay trials). No such experimenter error occurred for participant *33* (6 months old), so it is unclear why she looked to the correct side accurately in the test phase but not the training phase. Participant *33* did not, though, show a large difference in accuracy between the switch and the stay trials in either phase. Her accuracy during the test phase is closer to chance than all other participants' with the exception of *10* and *20* (52.6%), but her accuracy on switch trials in particular is 55.8% (stay trials = 48.61%).

Participant *42* (7 months old) also showed some ability to distinguish the two languages, though did not attend to the training phase the requisite amount of time and therefore should be excluded based on our original guidelines. He attended to the training phase only 24.7% of the time, but looked to the correct side 100% of that time. However, in the test phase, participant *42* attended 36.5% of the time and was looking to the correct side 68.4% of that time. Again, participant *42*'s results are ambiguous and should be interpreted with caution, but show tentative evidence of an ability to identify a code-switch and use that identification to anticipate the location of a re-appearing image. I will return to potential age effects in the discussion.

Participant *33* is the only participant who was assigned to the naturalistic condition and attended to the task for at least 33.3% of the training phase. Participant *42* was also assigned to the naturalistic condition but did not attend to at least 33.3% of the training phase, though he did attend to the test phase and look correctly during both phases. The remaining 7 participants with at least partially interpretable data were all assigned to the unnatural condition. Comparison between the two conditions is impossible and lacks meaning given these small, unbalanced sample sizes.

<u>Discussion</u>

In Experiment 4, 5- to 14-month-old English monolingual infants were trained to associate one side of a computer screen with French and the other side with Mandarin using an adaptation of the Anticipatory Eye Movement paradigm. Infants heard trials that were either presented all in one language or contained a code-switch, while moving images appeared and disappeared. The moving image was present at the beginning and end of each trial, but missing during the period of time immediately preceding and following the point that could contain a code-switch. I tested whether infants looked to the correct side of the screen after a code-switch (or matched mid-point) but before the object reappears, to determine if they could correctly predict where the object would re-appear. The original goal for this study was to compare this performance on switch trials and stay trials within each participant, and to compare participants assigned to the naturalistic and the unnatural stimulus conditions. However, due to a low yield rate and infants' general lack of engagement in the study even for those with scoreable data, such detailed comparisons are currently difficult or unreliable.

This study provides some preliminary evidence that the Anticipatory Eye Movement paradigm can be used to train infants between 5 and 14 months to categorize which language they are hearing, including to detect a switch between two languages. Of the six infants who attended to the study for at least 33.3% of the time range of analysis in both the training and the test phase, two of them looked to the correct side of the screen more than 50% of the time in both phases, two looked to the correct side of the screen more than 50% of the time in the training phase and approximately 50% of the time in the test phase, and two looked to the correct side of the screen more than 50% of the time in the test phase but less than 50% of the time in the training phase. No infant reliably attended to the screen during this region of analysis but looked in the *wrong* direction more than 50% of the time in both phases. However, with such a small number of participants with usable data, we cannot answer more detailed questions about how specific properties of the stimuli influenced infants' ability to detect a code-switch.

Potential age effects

There are tentative evidence for age effects in infants' ability to identify a switch between two unfamiliar languages. Namely, the five participants who show the best ability to correctly identify a language (> 33.3% attention and > 58% accuracy in the test phase) are between 6 and 8 months of age, while two participants who attended to the test phase for at least one third of the target range of analysis but looked to the correct side at chance levels (49.5% and 50.7%) were 11 and 12 months old. Disregarding accuracy, the six infants who attended to both phases at least 33.3% of the time have a mean age of 8.5 months, as compared to the 9.65 months of the total analyzable sample. This pattern suggests both that younger infants were better at attending to the screen and that younger infants *might* be better at discriminating unknown languages, although both findings should be interpreted with caution given the small sample size presented here. If such a pattern remains consistent with more participants such that it appears reliable, then several mechanisms may drive the observed relationship.

One possibility is that the greater proportion of attending to the screen and greater accuracy (especially during the test phase) among younger infants are independent processes, driven by separate patterns. Younger infants (six- to nine-months-old), who have head and neck control but limited independent mobility, are better at participating in the head-turn preference procedure than older infants, which similarly requires infants to sit on their parent's lap, listen to speech which they may or may not understand, and turn to look in either direction (E. K. Johnson & Zamuner, 2010). In the present paradigm, relatively young infants may be more content to stay seated either on their parent's lap or in a highchair for several minutes, look at a computer screen in front of them, and listen to a man tell children's stories, even if they do not know that man or recognize the language he is speaking, for similar reasons of limited mobility.

Infants' accuracy, which reflects a combination of their ability to detect a code-switch between French and Mandarin and to learn which side is associated with which language, could be independent of this willingness to sit and attend to the screen and instead driven primarily by their perceptual abilities. Infants in the 6-8 month range show a robust ability to discriminate between pairs of speech sounds which may not be meaningfully contrastive in the language(s) to which they are exposed (i.e. Werker & Tees, 1984, among others), while 11-12 month-old infants appear to be better at distinguishing phonemes in their native language and worse at distinguishing phones which are not phonemic in their language (Kuhl et al., 2006; Tsao et al., 2006).

Although discrimination of some of the specific consonant contrasts that may be useful in distinguishing French and Mandarin (i.e. Mandarin alveolopalatal/postalveolar affricates and alveolopalatal fricative vs. French postalveolar fricatives) has not been explicitly tested, work with English-learning and Mandarin-learning infants and adults suggests that younger English monolingual infants may have an advantage over older English monolingual infants at similar discrimination patterns. Tsao et al. (2006) found that English-learning 6- to 8-month old infants were better than 10- to 12-month-olds at discriminating the Mandarin alveolopalatal fricative ɕ from the Mandarin alveolopalatal affricate tɕ^h, though both groups were above chance. If this stronger ability to discriminate two non-native phonemes which differ in manner extends to a pairing of one native and one non-native phoneme which differ in manner and slightly differ in place (tɕ^h vs. ʃ[12]), then we might expect younger infants to be better than older infants at using consonant inventory to distinguish French and Mandarin. The presence of lexical tone is also a clear feature that distinguishes French from Mandarin; however, it is unclear whether we should expect an effect of age on English monolingual infants' ability to discriminate a speech segment that uses lexical tone from one that does not (Kalashnikova et al., 2024). Evidence is mixed regarding how non-tone-learning infants' discrimination of lexical tones changes with age, and the existing work primarily aims to assess the discrimination of two separate tones rather than speech samples that do or do not utilize lexical tone. If younger infants' relative strength in discriminating non-native phonemes does extend to contrasts that

[12] ʃ is used in French and English, but not Mandarin

cross Mandarin and French, then attunement to the native language may have led to older infants being less able to discriminate between the two languages using phonetic cues.

Such a benefit for younger infants should be stronger in the naturalistic condition than the unnatural condition. If the cross-language phonetic transfer which occurs in natural code-switching effectively “blurs” the difference between two languages, then infants who are less able to distinguish unfamiliar speech sounds should require a more abrupt change in phonetic detail at the point of the code-switch. Given the small sample size of meaningfully-interpretable data here, it is impossible to currently test that prediction.

An alternative possibility is that the measures of attention and accuracy are inherently related, in that both actually reflect an understanding of the task. To clarify, this attention measure is not attention over the entirety of the phase, but rather attention *during the periods of analysis*. Infants cannot be learning anything during the period of analysis in the training phase or the test phase, because the screen is blank. We would need to analyze the first and fourth segments of each trial, the time periods during which a visual stimulus is present, in order to assess the relationship between attention *during the segments that provide training* and later accuracy. Such an analysis is left for future research with a larger sample of participants. Instead, attention to the screen during these periods of analysis may reflect some level of understanding about how the task works; infants who expect a moving image to re-appear may be more likely to continue attending to the screen even when they cannot see anything, while infants who have not learned this basic structure may disengage as soon as the image disappears and only re-attend to the screen in response to a new visual stimulus. Further, correctly predicting the location of an upcoming visual stimulus may be rewarding and encourage participants to continue attending to the screen in future trials, and thus lead to additional learning and accuracy. In both hypotheses I assume that if infants could discriminate between French and Mandarin and detect a code-switch, they had that perceptual ability before starting the experiment. While the training phase may provide moderate training in distinguishing French and Mandarin, especially if the re-appearing images in the re-orientation segment of each trial are interpreted as "feedback" on the child's expectation, infants' perceptual sensitivity was already in place. The primary function of the training phase is in teaching infants to associate French with the left side and Mandarin with the right side, along with an experimental structure that includes a moving image which will re-appear.

If the first possibility (independent processes) is correct, then we would expect to see independent negative correlations between age and attention, and age and accuracy, in both the training and test phases, such that younger infants attend to the screen for a greater amount of time and are more accurate. However, there would not be a relationship between attention and accuracy. If the second possibility (attention drives accuracy) is correct, then we would expect a positive relationship between attention and accuracy in both the training and the test phases such that infants who attend to the screen for a greater proportion of time are more accurate, but that this relationship should be independent of age. There may also be a relationship between accuracy in the training phase and attention in the test phase, such that infants who successfully learned the relationship were more motivated to continue attending to the experiment through the test phase because they were continually rewarded by predicting the location of a moving image.

Table 10: Experiment 4: Correlation matrix between age, attention, accuracy for training and test phases

	Age	Training: attention	Training: accuracy	Test: attention	Test: accuracy
Age		0.10	0.10	0.22	-0.80
Training: attention			-0.21	0.47	-0.10
Training: accuracy				0.38	0.18
Test: attention					-0.13
Test: accuracy					

The above table shows a correlation matrix between age, attention in the training phase, accuracy in the training phase, attention in the test phase, and accuracy in the test phase for the 9 participants who attended to at least one phase for at least 33.3% of the time. This correlation matrix partially supports both hypotheses and suggests that investigating the relationship between age, attention, and accuracy may require a more sophisticated moderation analysis with additional participants. There is a small positive relationship between age, attention in the training phase, attention in the test phase, and accuracy in the training phase. However, the relationship between age and accuracy in the test phase is negative, suggesting that older infants were less accurate than younger infants. This relationship between age and test phase accuracy is in line with the age-as-an-independent-predictor hypothesis, but the relationship between age and attention in either phase is not. Disregarding age, infants who attended more to the crucial periods of analysis in the training phase also did so in the test phase, though they were not necessarily more accurate in either phase. Infants who were more *accurate* in the training phase both attended more to the test phase and were slightly more accurate in the test phase. However, attention in the test phase has a small negative relationship with accuracy in the test phase. The relationship between accuracy in the training phase and both attention and accuracy in the test phase is in line with the hypothesis suggesting dependence between attention and accuracy, though not all observed correlations support this hypothesis.

Low yield rate

One of the primary issues facing this study is an unexpectedly low yield rate. Attrition was caused by both video quality (including lighting) and child emotional state and attention. This study was run on Zoom instead of asynchronously in an attempt to pre-emptively address these concerns. In some cases, this attempt was successful. Most families did try to follow all set-up instructions from the experimenter, including turning lights on or off in the house to aid visibility of color changes in the screen-shared content reflected onto the child's face or re-scheduling the appointment for a time that the child might be more willing to sit still and participate. However, some were unable or unwilling to follow all instructions and exhibited some impatience with the set-up checks. Even for some of the participants who did comply with all set-up checks, the video was either partially or completely difficult to code. If the room in which they were participating was partially lit by sunlight, a change in cloud cover in the middle of the experiment could temporarily affect the visibility of color changes. Children who seemed content at the beginning of the experiment sometimes became fussy and upset partway through. Other families moved to a different part of the room or re-positioned their child or the webcam in the middle of the study, or were unexpectedly interrupted by other family members. These causes of the high attrition rate were rarely addressed once the study had started if the child was otherwise content and attending to the experiment, for fear of unnecessarily interrupting the child's learning of the language – direction contingency.

As discussed, some children became very fussy or upset partially through the study. In these cases, the experiment was often paused part-way through due to either parent request or experimenter judgment, so that the parent could attempt to soothe the child or remove a distractor from the environment. However, in no case in which the child was so fussy that pausing the experiment was necessary was the data ultimately usable. Other infants remained content but were disengaged in the task. These infants exhibited a variety of off-task behaviors including looking off-screen (outside windows; at individuals or objects in their homes; at their own hands or feet; turning to look at their parent), and reaching for or touching other objects surrounding them including toys, a bottle, and the computer. Such infants are generally excluded from detailed analysis due to not reaching established guidelines for proportion of time attending to the screen, even if the video was entirely scorable.

The exact cause of this disengagement among otherwise-content infants is unclear, though anecdotally it is likely the long period of time with no visual stimulus in the middle of every trial contributed to the problem. Several infants appeared to turn away from the screen when it went blank, but might re-attend when the visual stimulus re-appeared. Attention getters were presented periodically between trials, and continued playing until the experimenter judged that the infant was attending to the screen and reasonably content; attention was therefore high at the beginning of the trials that occurred immediately after an attention getter, but often decreased until the next attention getter. High rates of disengagement and fussiness have previously reported in similar paradigms; Albareda-Castellot et al. (2011) reported that to reach a sample size of 54 8-month-old infants, an additional 10 were tested but excluded due to fussiness or crying and 16 were excluded due to low attention during the time periods of analysis.

Discrepancy in trial length

As discussed, the analysis discussed here only includes trials whose coded length differed from the ground truth trial length by less than 200 ms. If more than 5 trials from a single participant differed in length between these two measures by more than 200ms even after review by the author, the participant was counted as “difficult to code” rather than “analyzed”; this was true for 6 participants. The remaining participants had an average of 1.18 trials which were excluded due to a trial length discrepancy (max = 5, n = 1). The origin of this discrepancy cannot be assessed for each individual trial, but likely resulted from both temporary fluctuations in lighting conditions in the participant’s home making it difficult to detect the color change and intermittent internet lag either on the part of screen-sharing the experiment or sending the video feed of the participant back to the experimenter over Zoom. Internet lag is only a potential cause of discrepancy for trials with a coded length greater than their ground-truth length, while temporary fluctuations in lighting could lead to trials being coded as shorter or longer than their true length.

Changes to a future study

The present study was designed to be run remotely over Zoom due to an expected difficulty in recruiting a sufficient number of infants in the necessary time frame to complete this dissertation. Indeed, it is unlikely that 44 infants would have participated in this study between August 2023 and April 2024 if it were instead run in person, given that as of the beginning of May 2024 the University of Maryland Infant and Child Studies database has a total of 26 infants who are willing to participate in in-person studies, were born full-term, hear English at least 80% of the time, and range in age from 5 to 23 months and thus may have been eligible to participate in the study at some point during the time when the study was active. This number includes participants who could have been added to the database after turning 15 months old and thus never would have actually been eligible. However, it is possible that if the study were run in-person instead of via Zoom, more participants would have resulted in ultimately usable data because we would have had full control over the environmental considerations including lighting and distractors. Rescheduling if the child was fussy or unexpectedly napping at the scheduled appointment time would have been more difficult though, and some children may have been more fussy and uncomfortable participating in an unfamiliar environment such as the research lab than they were in their own homes. Perhaps the number of meaningfully analyzable participants would have been the same if we had instead run the study in-person, with a smaller sample size but higher yield rate. However, without significant work in recruitment it is unlikely that the sample size would have been any larger using a different recruitment and data collection procedure. As the Infant and Child Studies consortium continues to re-build a database of potential participants after COVID-related disruptions and to find new ways to reach local families, a future in-person version of this study may find more success.

As discussed, an additional limitation of this study was the high number of infants who stopped attending to the study during the periods in which there was no visual stimulus. Although this aspect of the procedure was by design, the duration of the blank screen could have been slightly shorter by increasing the duration of the initiation segment and decreasing the duration of the preparation segment of each trial; this change might have led to marginal improvements in overall attention. I also could have used more engaging and salient visual stimuli, rather than colored circles and squares. Albareda-Castellot et al. (2011) used a character expected to be familiar to infants of the target age (Elmo); perhaps children would have been more motivated to look at the screen to find a similar character than they were to find a rotating circle. Future studies incorporating these design changes may be necessary in order to address the primary question of interest in Experiment 4, namely whether there is an effect of natural acoustic cues to code-switching on infants' ability to distinguish between two unfamiliar languages and identify a code-switch.

Chapter 5: Production approach

Introduction

Infant-directed speech (IDS), also sometimes known as "parentese" or "motherese", is a speech register often used when speaking to infants and young children. Acoustically it is broadly characterized by a higher and more variable pitch, lengthening of vowels, pauses, and other segments, and the use of specific intonational contours (see Soderstrom, 2007; Wang et al., 2018 for reviews). Vowels are often shown to be produced using an enhanced acoustic space in IDS as compared to adult-directed speech (ADS), though this result is not consistent (Englund & Behne, 2005; Kuhl et al., 1997). The content of IDS includes shorter and simpler sentences, as well as more repetitions resulting in a lower type/token ratio compared to speech to adults or even older children (i.e. Phillips, 1973; Soderstrom et al., 2008).

Despite these general properties, infant-directed speech is not a stable, unitary register that is consistent across or within individuals. It reflects a responsive interaction between two individuals, albeit an interaction in which one person drives communication (the adult) and the other provides occasional signals to understanding (the infant). IDS from the same parents differs when their child is physically present and responsive, as compared to when the infant is not providing direct feedback or is not present (Fernald & Simon, 1984; Jacobson et al., 1983; Smith & Trainor, 2008). Acoustic properties of infant-directed speech also differ based on the child's age, hearing abilities, and linguistic knowledge (i.e. Bergeson et al., 2006; Kitamura & Burnham, 2003; Narayan & McDermott, 2016, among others). These longitudinal and cross-sectional effects may reflect parents' belief about their child's knowledge and how they can affect their child's ability to understand speech in the moment and grow in their linguistic abilities. A recent meta-analysis indicated that pitch variability and vowel space area remain consistent in IDS up to approximately 36 months of age, while overall F0, articulation rate, and vowel duration in IDS may become more similar to ADS over the course of a child's first three years (Cox, Bergmann, et al., 2023). This work highlights both the importance of sampling across a wide range of ages and linguistic abilities and of considering the potential functions of infant-directed speech in relation to its form.

Individuals' cultural background and community norms may also influence the way they produce infant-directed speech. Cultures differ with respect to their belief regarding how infants and children should engage with the world, including the perceived importance of speaking directly to pre-verbal children (Keller, 2007). These belief systems, along with other aspects of community structure and expectations, result in differences in the amount of speech directed to infants across cultures (M. Casillas et al., 2020, 2021). Further, cross-cultural corpus studies suggest that even among cultures with similar beliefs about speaking to infants, acoustic properties of this infant-directed speech may differ. North American English shows greater differences in pitch between IDS and ADS than other languages, with potential exceptions of Danish and Swedish (Bernstein Ratner & Pye, 1984; Cox, Dideriksen, et al., 2023; Fernald et al., 1989), and may differ from other English dialects with respect to the instantiation of phonological rules in IDS as compared to ADS (Bernstein Ratner, 1984). Though little research has investigated properties of Spanish IDS in particular, some work suggests that vowels are produced in an expanded acoustic space relative to ADS (i.e. García-Sierra et al., 2021; Kalashnikova & Carreiras, 2022) and that some Spanish consonants may be produced differently in IDS than in ADS (Fish et al., 2017).

Role of infant-directed speech in learning

Infant-directed speech can serve a variety of functions, both social-affective and cognitive-linguistic. The pitch and prosodic characteristics of infant-directed speech may play a role in soothing infants and helping them to regulate negative emotions (Spinelli & Mesman, 2018). Cross-culturally and -linguistically, infants attend more to this speech that has a higher and more variable pitch, including when it is spoken in a language that is unfamiliar to them

(Byers-Heinlein et al., 2021; The ManyBabies Consortium et al., 2020). This attention-grabbing nature could lead to learning benefits for IDS, such that infants focus on the speech more and therefore have the capability to learn from it. Indeed, cortical tracking to IDS by infants is greater than that to ADS, which could aid learning (Kalashnikova et al., 2018; Nencheva & Lew-Williams, 2022). These prosodic differences between IDS and ADS, including variability in IDS prosody, can facilitate novel word learning, especially early in the lexical acquisition process (Graf Estes & Hurley, 2013; Ma et al., 2011).

Beyond the prosodic features of IDS, some research suggests that the register's acoustic properties at the segmental level may be specifically helpful for learning. Though this hypothesis has been primarily addressed in the domain of vowels, other segmental features such as the voice onset times of stop consonants have also been considered. We will begin by discussing the evidence from acoustics and learning outcomes in favor of hyper-articulation-driven learning, followed by evidence that opposes this belief.

Research in favor of vowel hyper-articulation and a learning benefit:

Early research on infant-directed speech provided strong evidence in favor of vowel exaggeration, by showing that the mean productions of the "point" vowels /a/ /i/ and /u/ are more distinct in IDS than in ADS (Kuhl et al., 1997). This effect is realized by shifting both the first and second formants such that the three vowels move apart from one another in the F1 x F2 multi-dimensional space. Exaggeration of vowels is associated with clear speech in general (Bradlow et al., 1996), and is not necessarily unique to infant-directed speech. Instead, it appears to be an adjustment that speakers commonly make in situations where speech may be more difficult to understand. This vowel exaggeration, often referred to as hyper-articulation or an increased vowel space area, has been attested in a variety of languages other than English including but not limited to Spanish (García-Sierra et al., 2021; Kalashnikova & Carreiras, 2022); Catalan (Kalashnikova & Carreiras, 2022); Swedish (Marklund et al., 2021); and

Warlpiri (Bundgaard-Nielsen et al., 2023). However, the effect is realized in a variety of ways cross-linguistically, as three different vowels can be shifted along the F1 and/or F2 dimensions in order to create a larger average vowel triangle. These changes also differ among individuals within a language community.

This natural variation in hyper-articulation in IDS allows for individual-differences comparisons of language development outcomes. Infants whose mothers produce IDS with a greater vowel space both show better discrimination of a consonant contrast (Liu et al., 2003), more phonetically complex vocalizations (Marklund et al., 2021), and greater vocabulary sizes and language proficiency levels (L. Dilley et al., 2020; Hartman et al., 2017; Kalashnikova & Burnham, 2018). Such patterns are also true in the context of bilingual language exposure: Spanish-Catalan bilingual infants whose mothers produce more vowel exaggeration in IDS show a more mature response to a non-native phoneme contrast (/b/ vs /p^h/) at nine months old, as do 11- to 14-month-old Spanish-English bilingual infants when listening to a native Spanish speech contrast (/d/ vs /t/) (García-Sierra et al., 2021; Kalashnikova & Carreiras, 2022). This pattern of responses may indicate greater perceptual attunement to their native languages among infants who hear more vowel exaggeration. However, most of these studies (with the exception of Hartman et al., 2017 which was part of a longitudinal study but included a small number of tokens) evaluated parental IDS and child behavior at the same time point. It is possible, then, that some properties of parental IDS are in response to what their child already knows, even if not at a conscious level, rather than to aid learning of currently unknown structures or patterns. Indeed, empirical research assessing whether infant-directed speech truly facilitates language *development*, especially at the morphosyntactic or lexico-syntactic levels, is limited (Kempe et al., 2024).

Evidence against vowel hyper-articulation or that it is not helpful for learning:

However, other work suggests that this expanded vowel space is not always present. Recall that as discussed previously, studies that show an overall effect of vowel hyper-articulation in IDS may have individual variability among parent participants. Furthermore, several studies do not show an overall effect of vowel hyper-articulation (i.e. Benders, 2013; Englund & Behne, 2005) and instead find either null effects or *hypo*-articulation in IDS. It is unclear whether such patterns are due to community- or language- effects, sampling methods, analysis method including the particular vowels measured, or recording practices including the naturalism of the prompt and instructions used to generate a speech sample. Regardless of the reasoning, this lack of consistency shows that vowel hyper-articulation may not be consistently available for infants; any theory of how the properties of infant-directed speech relate to learning must account for such variability across and within infants' experiences.

Even when it is present, this hyper-articulation does not necessarily aid infants' language development. Vowel space area is commonly measured using the three point vowels; an expanded vowel space may lead to increased distinctiveness among these three vowels. However, vowel changes in IDS may be associated with both a pulling apart of the mean F1 and F2 values of the vowels as well as additional variability within each vowel (Cristia & Seidl, 2014). This effect of increased variability could be particularly problematic for listeners' ability to distinguish interior vowels, which are not necessarily expanded to the same degree as point vowels (McMurray et al., 2013). Interior vowels, especially, may actually overlap *more* in IDS than in ADS due to this variability. Furthermore, formants are not the only quality of vowels that differs between IDS and ADS. For example, in Japanese, while vowels in spontaneous speech do use an expanded F1 x F2 space, their quality in terms of breathiness and variability also differs from carefully-read speech. When combining information from these dimensions, the acoustic distance between Japanese vowels is no greater in IDS than in ADS (Miyazawa et al., 2017). In fact, many phoneme contrasts may be clearer in Japanese ADS than in IDS (A. Martin et al., 2015).

While it is possible that within-category variability can aid learning by specifying the relevant features for defining a category (i.e. Lively et al., 1993), the utility of this variability to an infant is unclear. Behavioral and computational modeling results support acoustic findings suggesting that variability in IDS phonemes can make learning more difficult. Rosslund et al. (2022) found that increased pitch in Norwegian IDS (but *not* vowel space area) predicted larger vocabulary sizes for 18-month-olds, as did increased distinctiveness among all vowels. That is, parents who produced vowels with less internal variation, making them more distinct, had toddlers with larger vocabularies. A machine learning algorithm actually learned Japanese phonetic categories better using carefully-read speech or ADS than using IDS (Ludusan et al., 2021). Further, some researchers question whether infants truly learn phonetic categories as they are traditionally defined at all (Feldman et al., 2021; Matusevych et al., 2023; McMurray, 2023; Schatz et al., 2021). While we refer to phonetic categories throughout discussion of this study, the current analysis is actually agnostic as to whether infants develop phonetic categories or simply become attuned to relevant distributions of speech sounds in their language.

Consonants in monolingual infant-directed speech

Voice onset times of stop consonants also show a conflicting pattern of results regarding the presence, variability, and utility of hyper-articulation. A hyper-articulation account would predict that voice onset times are produced with a greater contrast in IDS as compared to ADS, such that voiceless phonemes have longer VOTs and voiced phonemes have shorter (and potentially more pre-voiced) VOTs. Such changes in both directions, so as to truly highlight the contrast between the two categories, are rarely found in the literature, especially when accounting for the differences in speaking rate between infant-directed and adult-directed speech.

Some studies find that the contrast between voiced and voiceless phonemes is increased in IDS relative to ADS (at least at some ages, if not all), which can be carried out either by increasing the VOT of voiceless stops or decreasing the VOT of voiced stops (Burnham et al., 2013; Malsheen, 1980). However, others find less distinction between the two categories in IDS (Benders et al., 2019; Rosslund et al., 2023; Synnestvedt et al., 2010), or an overall shift in VOT between IDS and ADS that affects the voiced and voiceless phonemes equally (Englund, 2005; McMurray et al., 2013; Økland, 2021; Sundberg, 2001; Sundberg & Lacerda, 1999). Rosslund et al., (2023) did find some evidence that VOTs in ADS and IDS became more similar as children aged from 6 to 12 months, and therefore gained linguistic knowledge and skill.

As discussed in McMurray et al. (2013) and Rosslund et al. (2023), many previous studies both used small sample sizes and did not account for speech rate in the measures of voice onset time. As VOTs are known to vary with speech rate, with slower speaking rate associated with longer VOTs for long-lag voiceless phonemes and increased pre-voicing for voiced phonemes, and IDS is often produced with a slower rate than ADS, accounting for this baseline difference therefore becomes important. Recall that McMurray et al. (2013) found that both voiced and voiceless stop consonants had longer VOTs in IDS than in ADS; this effect was driven by the difference in speech rate between the two registers. Rosslund et al. (2023) found similar effects in Norwegian, expanding this finding to a language that regularly uses pre-voicing to indicate voiced stops rather than short-lag/long-lag with aspiration contrast in English. One possible interpretation of these results is that the fine-grained acoustic features of infant-directed speech are not particularly designed in a manner that will simplify the learning process for infants. IDS may instead have other positive features, including those of affect and attention, which could indirectly support both language and social development. We will return later to this hypothesis and discuss it in the context of the present results.

Beyond vowels and the VOT of stop consonants, other phoneme characteristics and phonological processes may also be hyper-articulated or produced in a more canonical manner in IDS as compared to ADS (Bernstein Ratner, 1984). Indeed, both sibilants (Cristià, 2010) and stops (L. C. Dilley et al., 2014; Fritche et al., 2021) may be more distinct or stable in infant- or child-directed speech than in adult-directed speech. The English sibilants /s/ and /ʃ/ are produced in different places of articulation, which results in /s/ generally having a higher first moment (M1). This difference between the phonemes is particularly pronounced in speech to 12- to 14-month-old infants, as compared to speech to adults (Cristià, 2010). Several phonological processes and natural sources of variation in production can apply to the realization of English stops including assimilation with the place of articulation of a following consonant, glottalization, deletion, and being unreleased. Alveolar stop consonants /t/, /d/, and /n/ may be produced in their canonical forms at slightly greater rates in speech to children under two than in adult-directed speech (L. C. Dilley et al., 2014; Fritche et al., 2021). Other phonological rules that result in non-canonical realizations such as dental deletion, /ð/ deletion, and /ts/ -> /s/ are applied more frequently in speech to adults than speech to children with minimal expressive language (Bernstein Ratner, 1984), though some of these patterns may be specific to North American English. However, while mothers appear to exaggerate the vowel length cue to voicing, in which vowels before a voiced consonant have a longer duration than those before a voiceless consonant, to a greater extent in infant-directed than in adult-directed speech (Bernstein Ratner & Luberoff, 1984), such exaggeration may be at the expense of the realization of final consonants in some cases.

However, some conflicting evidence to this pattern of more canonical forms in IDS does exist. Khlystova et al. (2023) analyzed IDS and ADS corpora (from different parents) and found that while the canonical form was the most common allophone of the /d/, /s/, /z/, and /n/ phonemes in IDS, that pattern was not true for /t/. Further, these five phones, as produced in natural speech, were no more canonical in IDS than in ADS. Similarly, in elicited (read) speech, place assimilation (i.e. "greem beans" for "green beans"; "cap box" for "cat box") occurs just as often when directed to 18-month-olds as it does in similarly elicited speech to adults, and in a corpus of spontaneous infant-directed speech (Buckler et al., 2018). As with previously-discussed dimensions of IDS, it is likely that child age, participant culture, task demands, the feature under observation, and detailed methodological choices may all determine the extent to which any difference between IDS and ADS is in a direction that *a priori* appears to be facilitatory or inhibitory for the development of speech sound representations.

It should be noted that in all of these cases, speech to infants and children does not perfectly distinguish between phonemes at the phonetic level or apply optional phonological rules; variability in speech production exists regardless of the person to whom the speech is directed, and as discussed, some variability may be necessary for learning. Children hear both canonical and non-canonical forms of phonemes regularly. However, as discussed by Bernstein Ratner (1984), the local context of phonological rule application in infant-directed speech may be particularly important for children's comprehension and ability to learn particular sounds. That is, in that study the mothers would sometimes produce an utterance using a phonological rule that results in a non-canonical form, observe that their child did not appear to understand, and then restate the sentence without that phonological rule, thus producing a more canonical form. Such a pattern of behavior may both increase understanding in the moment and provide the child with the input needed to generalize a pattern of when particular rules may be applied.

Somewhat similar patterns are shown in a computational modeling approach from Adriaans & Swingley (2017): models trained on prosodically exaggerated vowels in IDS were better at later categorizing untrained tokens than models which were trained on nonexaggerated vowels. This finding, especially combined with infants' propensity to listen to prosodically interesting speech, suggests that local changes in speech production could be paired with broader content or contextual cues which attract attention and focus in a way that aids learning. As further support of this claim, Hitczenko & Feldman (2022) show that tracking the context in which a particular token occurs along with the properties of that token itself can aid a computational model's ability to categorize speech sounds.

It may be that the linguistic and extra-linguistic situations in which mothers produce particular tokens both influence that token's relative weight in infants' overall cognitive linguistic representations and provide additional information about how it should be interpreted. Using code-switching as an example, tokens produced in the presence of code-switching might be weighted relatively highly (i.e. if such language switching increases attention or requires additional cognitive effort to process) or relatively weakly (i.e. if code-switching results in general confusion). Independent of this information-weighting question, tokens might be produced differently in the presence of code-switching than in a fully unilingual environment. A mechanism whereby infants use information about the linguistic and extra-linguistic situation to interpret a given token (phoneme, word, etc.) is only useful if there is some dimension (i.e. the presence of code-switching) that can predict a form's realization. The present study is not designed to address the first part of this broader proposal (whether code-switching might impact how infants and young children weight information). However, it does aim to take a step toward addressing the second part by evaluating language mode and code-switching as potential contextual dimensions along which voice onset time may vary in infant-directed speech. We evaluate both global context in terms of language mode (unilingual vs. code-switching) and register (adult-directed vs infant-directed) and local context in terms of utterance type (unilingual vs code-switching) and switch type (English-Spanish vs. Spanish-English; distance to that code-switch).

Infant-directed speech from bilingual parents

The detailed properties of infant-directed speech as produced by bilingual parents is highly understudied, which inspired this analysis. However, two existing studies describe acoustic aspects of this speech by Spanish-English (Fish et al., 2017) and French-English (Danielson et al., 2014) parents. These studies show that in unilingual contexts the two languages are produced differently from one another in a way that reflects previously-described separation at the phonological level, though cross-linguistic interaction does occur in both speech to infants and speech to adults. However, some evidence suggests a greater level of distinction between the two languages in infant-directed as compared to adult-directed speech.

In a study of French monolingual, English monolingual, and French-English bilingual mothers of 11-month-old infants, Danielson et al. (2014) found that the two languages differ on the properties of fundamental frequency, vowel duration, and vowel height when spoken by both monolinguals and bilinguals, though patterns may differ slightly for vowel backness. Parents were given a set of labeled objects and instructed to play with their child while describing the objects for several minutes. Bilingual parents did this twice (once in each language), while monolingual parents completed the task once. The authors then identified all tokens of three point vowels (/a/ /i/ and /u/), and measured F0, F1, F2, and duration for all tokens.

One particular point to consider with respect to this analysis is that only nine bilingual mothers participated, while 20 English monolingual and 22 French monolingual mothers participated, and that the bilingual sample may be more variable simply due to relatively low statistical power. Further, all bilinguals were determined to be highly balanced in English and French on the basis of proficiency, frequency of use, and foreign accentedness as rated by native speakers. The patterns from this bilingual sample therefore may not necessarily be generalizable to a bilingual mother who learned one of the two languages later in life and is relatively dominant in her first-learned language.

The analysis itself, likewise, is limited in the conclusions that it can draw. Languages (English vs French) were compared for the monolingual groups and within the bilingual group, with a goal of determining the extent to which a single bilingual distinguishes the two languages similarly to monolingual speakers of each language. However, monolinguals and bilinguals were not compared within productions of the same language. Despite these limitations, the authors found evidence of differentiation between English and French among bilingual speakers when talking to their children. This differentiation in pitch, vowel duration, and vowel height and backness (for some vowels) may help the bilingual infants to discriminate their languages.

While Danielson et al. (2014) only studied infant-directed contexts, and therefore was unable to determine whether the observed patterns are consistent properties of their speakers' representations and articulatory gesture patterns in each language or specific to the experience of directing young language learners, Fish et al. (2017) did compare infant-directed and adult-directed speech. English monolingual and Spanish-English bilingual parents of 11- and 14-month-olds recorded a series of words beginning with stop consonants to their infant and to another adult, both as single words and in a sentence context. The authors measured the voice onset time of voiced and voiceless stops in both languages, and tested how this latency was affected by phonemic voicing category, language, participant group, register, and context. While both monolinguals and bilinguals produced overall longer VOTs in IDS than in ADS, monolingual parents' English VOTs diverged between the voiced and voiceless categories more when they were talking to an infant than when they were talking to an adult, perhaps in an attempt to pull apart the two phoneme categories within the language. Bilingual parents produced more pre-voicing of phonemically voiced consonants in Spanish in IDS than in ADS, but did not differ in their differentiation of the two English stop categories across the two speech registers.

This pattern of increased pre-voicing for voiced consonants and longer VOTs for voiceless aspirated consonants exhibits hyper-articulation of the voicing contrast in opposite directions for the two participant groups. Spanish-dominant bilingual parents highlighted a change in an acoustic direction which fits Spanish norms but not English norms, while English monolingual parents did so in a direction which fits English norms but not Spanish norms. The results therefore provide some evidence that parents' relative distribution of voice onset times in voice and voiceless stops differs when talking to their infants as compared to adults, in a way that may benefit learning by increasing the separation of these two categories. However, this increased separation may only occur in the adult speaker's first-learned language, as the bilingual parents showed this pattern in Spanish but not English. Furthermore, the study did not address the separation of the two languages among the bilingual parents, nor the extent to which this separation differed between IDS and ADS.

An important consideration is that both of these studies only considered unilingual contexts: bilingual parents were informed which language to speak during a particular interaction, rather than having the freedom to switch between languages if they so desired. As established in the overall introduction to this dissertation, some degree of code-switching is common in speech to children from bilingual adults. It is therefore possible that in addition to the inherent constraints in speech production due to dictating specific words for participants to use or sentences to say, these analysis methods failed to assess the diversity of bilingual parents' language behaviors by focusing on speech in a unilingual mode.

<u>Acoustics of code-switching and methods of analysis</u>

Though the research regarding the acoustic properties of bilingual infant-directed speech has not previously considered code-switching, this question is well-represented in the adult-directed speech literature. As discussed in the introduction for the production aim of this dissertation, several studies using both natural speech and cued speech suggests that code-switching may have a local effect on the production of particular sounds. Speakers appear to produce English voiceless stops that are minimally removed from Spanish with a shorter, and therefore more Spanish-like, voice onset time. However, the exact pattern of this result may vary across participants' language backgrounds, task, or details of the analysis method. Here we will focus on the methodological details of two corpus-based analyses of the properties of Spanish-English adult-directed code-switching: Fricke et al. (2016) and Balukas and Koops (2015).

Recall that Fricke et al. (2016) found that English voice onset times were produced slightly shorter, and therefore in a more Spanish-like manner, in anticipation of a code-switch from English into Spanish. Crucially, there was an additional effect of distance to the code-switch, such that VOTs directly before that switch into Spanish were shorter than those that occurred six seconds before the code-switch. In order to assess this effect, Fricke et al. (2016) considered only tokens of voiceless stops that were produced in a code-switched utterance. Utterances were defined in the original corpus such that each contained one main clause (Deuchar et al., 2014); target phonemes were never more than 6.4 seconds from a code-switch within an utterance. Fricke et al. evaluated whether the code-switch in that utterance was an English-to-Spanish switch, such that the target English token preceded Spanish, or a Spanish-to-English switch such that the target English token followed Spanish. These two switch directions were analyzed separately, with separate models fit for the English-to-Spanish and the Spanish-to-English switches. While a significant effect of distance was found for the English-to-Spanish direction with greater distances associated with longer (more English-like) VOTs, this was not true for the Spanish-to-English direction. In all analyses, Fricke et al. controlled for the consonant (/ptk/), length of the target word in syllables, and speech rate operationalized as average syllable duration over the entire utterance, as well as a random effects structure which included random intercepts for speaker, conversation, and target word; and by-speaker and by-word random slopes for the main predictors.

Balukas & Koops (2015), meanwhile, found an effect of distance to the nearest code-switch on English VOTs such that tokens closer to a code-switch were produced with a shorter VOT, but only assessed those tokens which *followed* a code-switch. The corpus used for these analyses (NMSEB; Cacoullos & Travis, 2018; Travis & Cacoullos, 2013) separated all utterances according to their intonational unit, which are tagged according to the language from which the lexical content is derived (English-only, Spanish-only, both). Two-thirds of their English tokens were produced within five seconds following a Spanish-to-English code-switch, an additional 18% were produced within 15 seconds following a code-switch. The greatest relationship between distance from a code-switch and VOT was for tokens that followed a code-switch by only a few seconds, as assessed by log-transforming the distance variable. Balukas & Koops normalized for speech rate and word length by including the length of the following vowel and the length of the target word in syllables as covariates in the model. Their random effects structure included random intercepts for speaker, word, and phoneme, as well as vowel height and approximant type, but no random slopes.

Two main conclusions can be drawn from this comparison of the methodological approaches taken by Fricke & colleagues and Balukas & Koops. First, the effect of distance to a code-switch on VOT appears to be local, and to occur within five seconds of that code-switch. It is unclear if this pattern is due to a genuine temporally short influence of code-switching on VOT, if it reflects a prosodically driven effect such that code-switching only influences VOTs within a prosodic unit (which temporally close items tend to be), or is an artifact driven by the fact that code-switching is frequent in the previously-analyzed corpora thus leading English /ptk/ tokens to often appear in close proximity to a code-switch. In the present study we will address this effect of proximity and the degree to which effects of code-switching are temporally constrained using separate models that evaluate only those tokens that occur within three seconds of a code-switch, as well as all tokens produced in a story in which code-switching was licit. The second conclusion is actually more of a question; it is unclear whether the effect of code-switching on English VOT can be better described as anticipatory, as in Fricke et al., or perseverative, as in Balukas & Koops. In the present study we will assess each direction separately, similarly to Fricke et al., and will further test if the effect of distance to a code-switch in each direction interacts with speech register (ADS vs. IDS).

The present study

The current study aims to fill a gap in the literature by evaluating naturally-produced infant-directed and adult-directed speech from the same individuals across three language modes: English-only, Spanish-only, and code-switching. This design allows us to first assess the roles of language mode (unilingual vs. language switching) and register (infant-directed vs. adult-directed), and to then assess the acoustic-phonetic properties of code-switched speech in particular. This study will serve several related goals. First, it will contribute to the broader literature regarding the pedagogical implications of infant-directed speech by evaluating how parents separate their languages in acoustic space when speaking to infants. It will also add to our understanding of language control at the phonetic and phonological levels during code-switching, and the potential role of speech register on that language control. In a broader sense, though, this work will describe the nature of code-switched speech directed to infants and directly compare that speech to unilingual contexts. This analysis will allow us to better understand the properties of the input that infants must be learning from. Our models of infant language acquisition and representation cannot be accurate if they do not reflect the true structure of the input available to infants (see Lavechin et al., 2024 for a discussion of consequences of input that is not ecologically valid in computational models). By establishing the properties of this input, and particularly the extent to which two languages may be separated at the phonetic level in infant-directed speech, we will be better able to evaluate infants' requisite perceptual and segmentation abilities.

Methods

Participants

This study was conducted remotely and asynchronously in order to allow us to recruit from beyond the local university community. Participants were recruited via several sources including Prolific, Children Helping Science, social media, local pediatrician offices, daycares, churches, community centers, and word of mouth. The current sample size for this study is 10 participants, 9 mothers and 1 grandmother, who are all proficient in both Spanish and English (M_{age} = 36; all female). Four participants are English-dominant, three are Spanish-dominant, and three are relatively balanced in their language dominance and proficiency.

Table 11: Production study: Participant demographic information

Participant	***Participant age (years)***	***Child Age (months)***	***Language dominance category***	***Reported % of time code-switching in speech to child***
02	31	15	Balanced	10
07	38	18	English-dominant	20
08	24	12	English-dominant	5
14	33	6	English-dominant	100
15	32	20	English-dominant	5
28a	30	6	Spanish-dominant	0
28b	31	6	Spanish-dominant	100
29	69	12	Spanish-dominant	100
31	36	14	Balanced	40

41	36	9	Balanced	5

The study was approved by the University of Maryland Institutional Review Board, and all participants provided a statement of informed consent. Participants received a $20 online gift card upon completion of the study; it was designed to take approximately one hour to complete though some participants needed more or less time depending on the level of detail included in the stories they told. Participants were encouraged to break up their participation over the course of several days and to avoid telling all stories back-to-back, but were not given specific instructions regarding how to do this. Some participants completed the study within a few days following providing informed consent while others needed frequent reminders over the course of multiple months.

Procedure

All participants completed a questionnaire designed to assess their use of Spanish and English in several situations including when talking to other adults, when speaking to their infant, and when in specific interactional environments (reading books, telling personal stories, telling family stories, telling cultural stories, telling make-believe stories). The questionnaire also asked about how often they engage in several characteristics of infant-directed speech, including using shorter sentences, simple words, repeating words, holding out particular sounds, speaking slowly, and speaking in a "sing-songy voice".

Participants recorded themselves narrating six wordless picture books. Each book contained 30 pages of illustrations and featured characters, items, places, and actions containing the target phonemes (/p/, /t/, /k/) in the language(s) the participant was to speak in while narrating that story. Two books were designed to be told entirely in English, two entirely in Spanish, and two while freely switching between the two languages. In those two code-switching stories, book titles were provided in both languages (i.e. Carmen's Dream / El Sueño de Carmen) and items that appear in the book contain the target phonemes in both languages (i.e. clown/payaso, kangaroo/canguro, crocodile/cocodrilo, cake/pastel, tea/té, castle/castillo). In each language context, participants told one story to their infant (IDS register) and one to another Spanish-English bilingual adult (ADS register). We required that participants speak directly to their infant, rather than simply tell the story as if they were speaking to an infant, as previous work has found that some properties of infant-directed speech are driven by the contingent social interaction between adult and infant (Braarud & Stormark, 2008; Fernald & Simon, 1984; Smith & Trainor, 2008). For the adult-directed speech register we asked participants to talk to any adult in their life who knew both English and Spanish, which could be a spouse, other family member, friend, neighbor, etc.

Figure 13 shows a representative page of the English IDS book. The characters (sisters named Calli and Paula) are labeled, and we see them in their kitchen carrying out the steps to make a pizza. In other books, a girl goes for a walk around her city, a girl has a vivid dream, a pair of brothers play basketball, some animals play with a ball together, and a dog befriends a turtle. All books are available at https://osf.io/bkjga/; data including audio recordings and transcriptions will be made available on CHILDES from consenting participants. Following recording, participants uploaded their audio files to a secure form to be accessed by the research team.

Figure 13: Production study: Example page from the English infant-directed storybook

Participants received all books as PDFs sent over email and were told to look at the pictures on their device or to print them and look at the print copies. Participants were also told that if requested, the research team would print all six books and mail them to the participant's home; one individual did request this service but to date has not completed the recordings. We did not ask participants to report how they viewed the stories while recording. The turning of pages is audible in the recordings from several participants indicating that they printed the books, while another parent can be heard telling her child "no no don't touch the computer."

Participants were also provided with instructions to create high-quality audio recordings. These instructions included information about minimizing background noise (i.e. turning off the tv and other loud electronic devices, moving away from sources of mechanical noise such as an air conditioning unit, encouraging other people or animals in the home to be quiet while recording, not recording at a time when the infant was unusually fussy or vocal), for ensuring that the participant's voice was clearly audible and not subject to excessive reverberation (i.e. being physically located in a room with soft surfaces and speaking toward those surfaces, keeping the recording device close to the participant's mouth), and for recording at a high quality (using freely-downloadable but better than the default voice recording apps, adjusting app settings appropriately). We have no way to know for sure the extent to which participants followed these guidelines. However, in most cases there was little obtrusive background noise or evidence of reverberation, and any conflicting noise arose from the infant's vocalizations.

Analysis

Data transcription, annotation, and phonetic analysis were conducted using Praat (Boersma & Wernick, 2023). All utterances were segmented and orthographically transcribed by a native speaker of the language(s) being spoken. The Montreal Forced Aligner (McAuliffe et al., 2017) was then used to approximately annotate word and phoneme boundaries. All voiceless stops in syllable-initial position were identified and their VOT boundaries were hand-corrected by either the first author or a trained research assistant. VOT was measured from the stop release burst to the onset of periodicity associated with the following vowel (or voiced approximant, in the case of words such as "play" or "cry"). In some cases, VOTs could not be measured because background noise (including the child's voice) overlapped with the stop release burst and/or the onset of periodicity, the participant was whispering, the phoneme was produced with pre-voicing, or the phoneme was produced as an affricate (in the case of words such as "tree"). Such tokens were excluded from analysis. In all cases the boundaries of the vowel or voiced approximant immediately following the target phoneme were also hand-adjusted. The length of this following phoneme, as well as its first, second, and third formants, were measured using Praat.

Several additional annotations and measurements were made for the two code-switching stories from each participant in order to identify the nearest code-switch from each token. The boundaries when the participant spoke each language were identified. For cognates or proper nouns that are potentially language-ambiguous, a combination of the surrounding language context and the participant's pronunciation was used to assign the word to a language when possible. For example, some characters in the code-switching stories had names which would be produced differently in English and Spanish (Thomas/Tomás, Peter/Pedro). When such names were produced in a Spanish-like manner they were tagged as Spanish, even if the surrounding language context was English. For names such as "Carmen" whose pronunciation does not necessarily differ between English and Spanish, we relied on the surrounding language and assumed that the participant had not produced an insertional code-switch for the character name. If a language-ambiguous cognate, proper name, interjection, or onomatopoeia was produced outside of a sentence context, its language was un-tagged and the token was excluded from analysis.

For each token I then measured the distance between the offset of the nearest preceding "other language" and the release of the stop burst associated with the beginning of the stop phoneme production, and the onset of periodicity associated with the end of the stop phoneme production and the onset of the nearest following "other language". If both of these distances to the other language were greater than 3 seconds, the token was considered to be from a unilingual utterance. Otherwise, it was categorized as a code-switched utterance and the shorter distance was taken to be the relevant distance for the purposes of potential effects of cross-language transfer. As discussed above, Fricke et al. and Balukas & Koops relied on previously-coded utterances for their analyses of naturalistic adult-directed speech. Infant-directed speech often contains short utterances which consist of a small number of words, which is atypical in ADS. Such short utterances were seen regularly in the data, as were long utterances in which a participant told a large segment of the story without a clear prosodic boundary. In order to avoid a situation in which tokens which occur temporally proximate to a code-switch but are separated by a prosodic boundary are excluded from analysis while tokens which are separated from a code-switch by up to 20 seconds are included, this consistent temporal distance was instead used as a marker of whether a token should be counted in the same "utterance" as a code-switch. However, it should be noted that this categorization is purely temporal in nature and is not based on the structure of prosodic units in the speech.

Each target phoneme was than categorized as being most affected by a Spanish-to-English or English-to-Spanish code-switch, based on the relative position of the two languages. For English target phonemes, Spanish-English switches could lead to possible perseverative effects of Spanish on the English production while English-Spanish switches could lead to possible anticipatory effects of Spanish on the English production. The reverse is true for Spanish target phonemes, such that Spanish-English switches would have anticipatory effects and English-Spanish switches would have perseverative effects of English on production of the target Spanish.

To give an example, take a participant who spoke Spanish from 1.34 to 5.86 seconds, English from 6.46 to 9.74 seconds, and Spanish from 10.21 to 19.38 seconds, including a target phoneme in English with a 55ms VOT from 7.534 to 7.589 seconds. The breaks between these language segments, from 5.86 to 6.46 seconds and from 9.74 to 10.21 seconds, represents either a pause in the participant's speech (in some cases the infant vocalized during such pauses, while some pauses are periods of silence) or a time when the participant produced language-ambiguous speech. This language-ambiguous speech, including onomatopoeia ("bumpadibump") and exclamation ("ah", "no", "woo"), was excluded from analysis when it occurred between two segments which clearly belong to different languages. The distance to Spanish preceding this token is from 5.86 to 7.534 seconds, for a distance of 1.674 seconds. The distance to Spanish following this English token is from 7.589 to 10.21 seconds, for a distance of 2.621 seconds. At least one of these distances is less than 3 seconds and the shorter distance here is the first one, the distance from preceding Spanish to the target phoneme. Thus, this token would be categorized as coming from a code-switched utterance. Specifically, production of this phoneme is likely most affected by a Spanish-English switch where any influences of Spanish on this English production would be perseverative in nature, and that nearest Spanish is 1.674 seconds away. A total of 66 tokens (out of 355) fall within three seconds of Spanish in both directions and are thus potentially ambiguous as to whether the effect is perseverative or anticipatory, though they are only included in one analysis, while the remaining 289 tokens occur within three seconds of Spanish in only one direction.

For each target phoneme, the VOT was divided by the length of the following phoneme to create a consonant-vowel ratio, following the same technique as McMurray, et al. (2013). As voice onset times are affected by speech rate and word length (Kessinger & Blumstein, 1997; Miller et al., 2009), it is important to ensure that any effects we see of infant-directed speech are truly register-based changes to voice onset times, rather than unintentional byproducts of speech rate modulation. In most cases this following phoneme was a vowel, however for some words it was actually a liquid (/l/ or /ɹ/).

In addition to these measurements of the voice onset times of each voiceless stop phoneme, the average fundamental frequency (F0) of each utterance was calculated using Praat Pitch tools. As described above, utterances for the present purposes are partially based on language choice. That is, if a participant produced a code-switch as a single-word insertion, that insertion would both be labeled in a different language, and would be segmented into a separate utterance, while the remainder of the sentence surrounding this insertion would be broken into two utterances. Therefore, some utterances are quite short. When a participant used a single language for a longer stretch of time, utterances boundaries were set at prosodic or syntactic end-points. Pitch was measured over each file in its entirety, using a time step of 0.015 seconds, a pitch floor of 50 Hz, and a pitch ceiling of 600 Hz. The average pitch of each utterance was then calculated.

Results

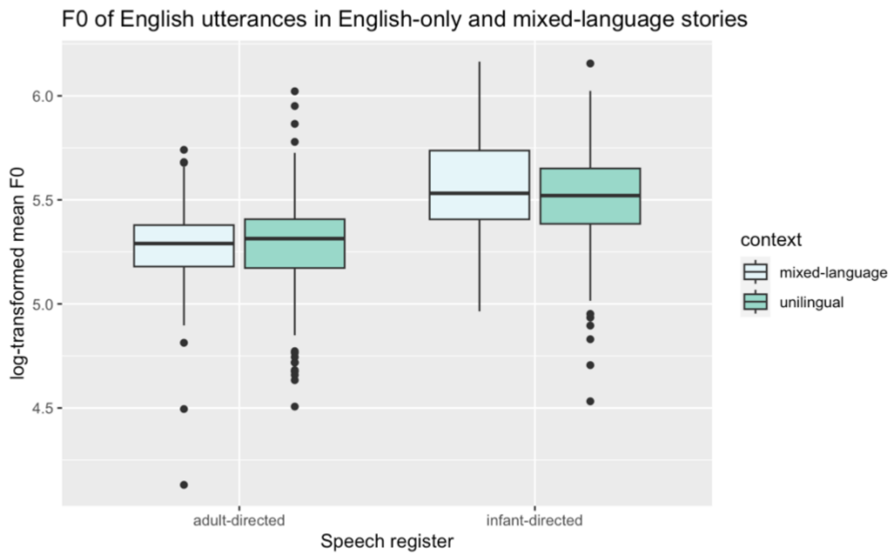
All described analyses focus on productions of English. Transcription and scoring of productions of Spanish are currently in progress. As previous studies of phonetic transfer of voice onset time in English-Spanish code-switching suggest that English voice onset times may be more susceptible than Spanish voice onset times to this transfer, the initial analysis is targeted at English.

For all analyses I started by attempting to fit a model which included all random effects which came from design and sampling decisions; however, such a maximal model failed to converge in all cases. Random effects were removed starting with those which were the most complex or least relevant to the primary research question. For analyses which should be directly compared (i.e. testing the same effects on two different dependent variables), results are based on the maximal random effects structure that converged for both models, even if one model could be fit using additional random intercepts. This approach allows for the most accurate comparison, even if it is not necessarily the most valid for each individual analysis.

Analysis 1: Pitch

The goal of Analysis 1 is to test if participants produced perhaps the most robust characteristic of infant-directed speech: higher and more variable pitch in speech to infants. Initial inspection of the pitch of each utterance suggests that the distribution is somewhat positively skewed (skewness = 0.87). This non-normality is addressed by taking the natural log of the pitch of each utterance; the distribution is log-normal (skewness = -0.08). The natural log of the pitch in Hertz is used as the dependent variable in this analysis, though patterns are consistent when the raw pitch in Hertz is used instead.

Figure 14: Production study - Analysis 1: Log-transformed mean F0 of English utterances in English-only and mixed-language stories



A linear mixed-effects model was fit to test the interaction between register (infant-directed speech vs. adult-directed speech) and story type (English-only vs code-switching) on the natural log of the average F0 *of English*. The maximal convergent model for this analysis included a random intercept for each participant and random slopes of both register and story type for each participant. This analysis suggested a main effect of register, such that infant-directed speech is higher-pitched than adult-directed speech ($t(9.1) = -8.0$, $\beta = -0.126$, $p < 0.001$) and an interaction between register and story type ($t(1205) = -3.13$, $\beta = -0.015$, $p < 0.01$), but not a significant main effect of story type ($p = 0.46$). This interaction between register and story type was explored by analyzing the effect of story type separately for adult-directed and infant-directed speech.

Separate linear mixed-effects models were fit testing the effect of story type in each register, using random intercepts for each participant and a random slope of story type for each participant. Neither model suggested a significant effect of story type (adult-directed: $t(8.40) = -0.346$, $\beta = -0.006$, $p = 0.738$; infant-directed: $t(9.48) = 1.391$, $\beta = 0.026$, $p = 0.196$). However, numerically it appears that pitch is higher in the code-switching than the unilingual stories in infant-directed speech, but lower in the code-switching than the unilingual stories in adult-directed speech. Given the small sample size, it is unclear the extent to which this significant interaction should truly be interpreted. I will return to this question in the discussion.

I also addressed the common finding of more pitch *variability* in infant-directed as compared to adult-directed speech using Levene's test for homogeneity of variance across groups. Two Levene's tests were conducted: one to test homogeneity of variance between the infant-directed and adult-directed stories, and one to test homogeneity of variance between the unilingual and code-switching stories. These tests suggest that pitch differed in its variability between the two speech registers, such that infant-directed speech was more variable than adult-directed speech ($F(1) = 14.408$, $p < 0.001$), but did not differ in its overall variability between the two story types ($F(1) = 0.816$, $p = 0.37$).

As the current study was not designed to assess pitch, pitch differences between infant-directed and adult-directed speech, including intonation contours, are not further investigated in the present set of analyses. Analysis 1 establishes that participants did indeed produce at least some of the acoustic characteristics associated with infant-directed speech when telling stories to their babies.

<u>Analysis 2: Global effect of code-switching on VOT (story type; English-only vs. code-switching)</u>

In the following analyses, I assess how register and language context influences the voice onset times of English voiceless stops, starting at a global level and ending at local effects. In all analyses, I assess both voice onset time (in ms) and the ratio between voice onset time and the length of the following vowel. However, both distributions are positively skewed. I therefore use the natural log of both dependent variables for our analyses, which reduces the skewness from 1.36 to -0.70 (for VOT in ms) and from 2.42 to -0.35 (for consonant/vowel ratio). While VOT remains somewhat negatively skewed, further transformations may lead to overfitting of the data, and the statistical analysis methods used are somewhat robust to non-normality (Schielzeth et al., 2020).

The analysis of the global effect of code-switching included all tokens of English voiceless stops, including those produced in English-only and code-switching stories. A total of 1941 tokens were included in this analysis, across all 10 participants and 3 phonemes: 665 from code-switching stories and 1276 from English-only stories. Such a discrepancy in number of tokens between story types is expected, as only English tokens are assessed in this analysis. Participants' speech was split between English and Spanish in the code-switching stories, but they did not tell stories that were twice as long. These 1941 tokens are evenly split between adult-directed and infant-directed speech.

The effects of register (infant-directed vs. adult-directed), story type (English-only vs. code-switching) and the interaction between these two factors were assessed, while controlling for word length. Register and story type were effects-coded (-1 and 1). Word length was positively skewed, so all values were expressed in terms of milliseconds and then log-transformed. The maximal random effects structure which converged for the analyses of VOT and consonant-vowel ratio included random intercepts for both phoneme and participant, random slopes of register for each phoneme, and random slopes of register and story type for each participant.

Figure 15: Production study - Analysis 2: English VOT (ms) by register, story type

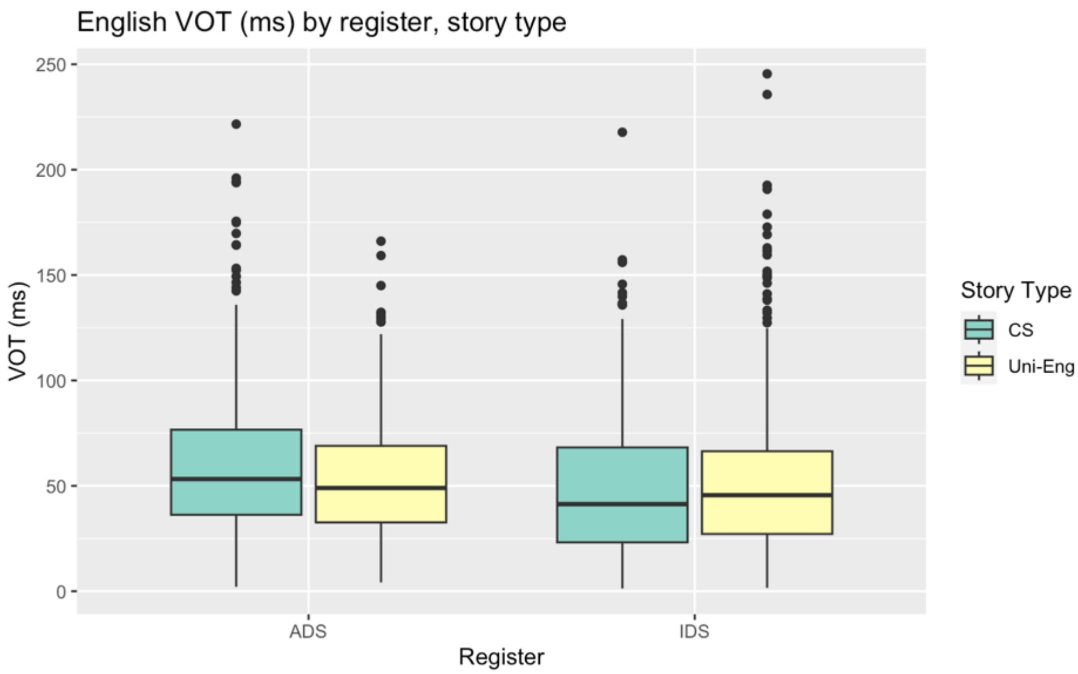


Figure 16: Production study - Analysis 2: English VOT (ms) by register, story type – log10 transformed

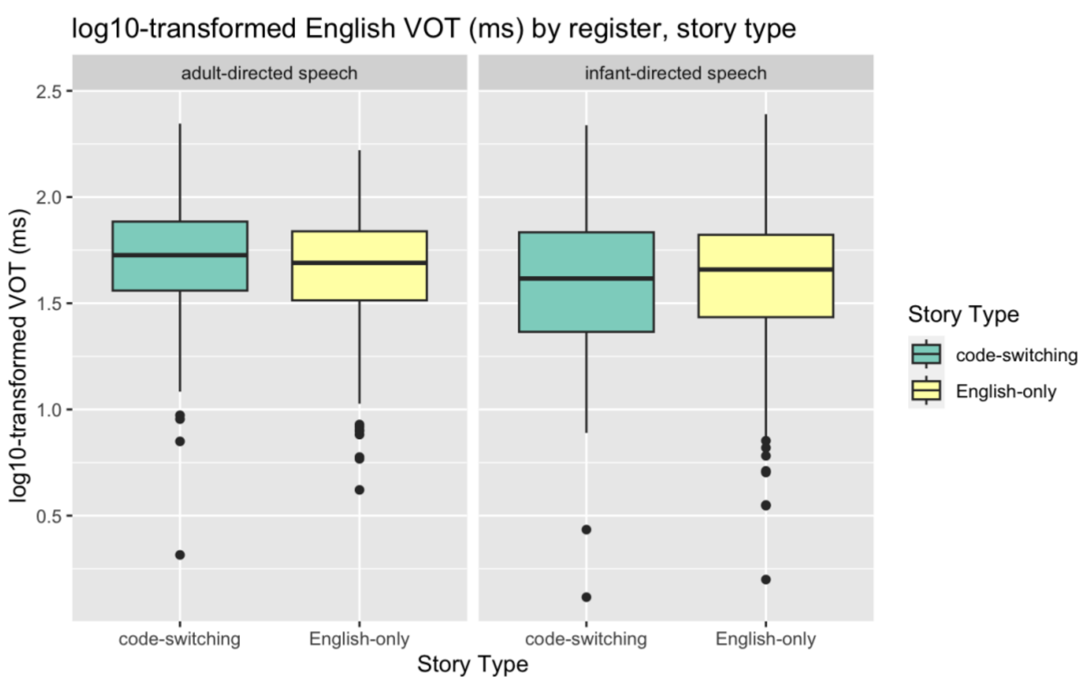


Figure 17: Production study - Analysis 2: English consonant-vowel ratio by register, story type

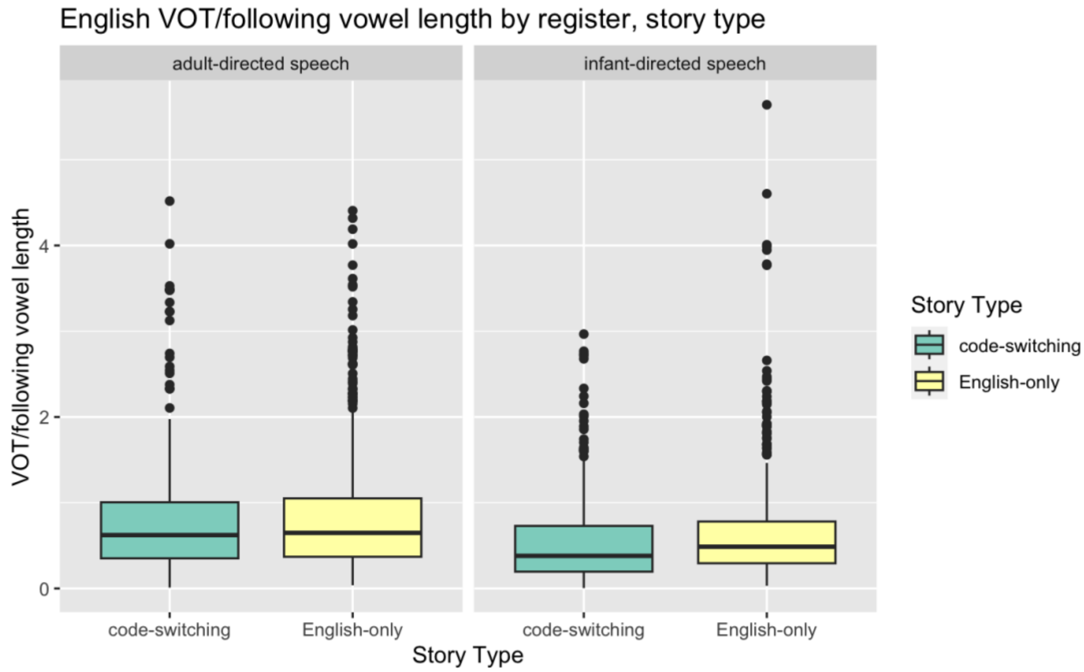
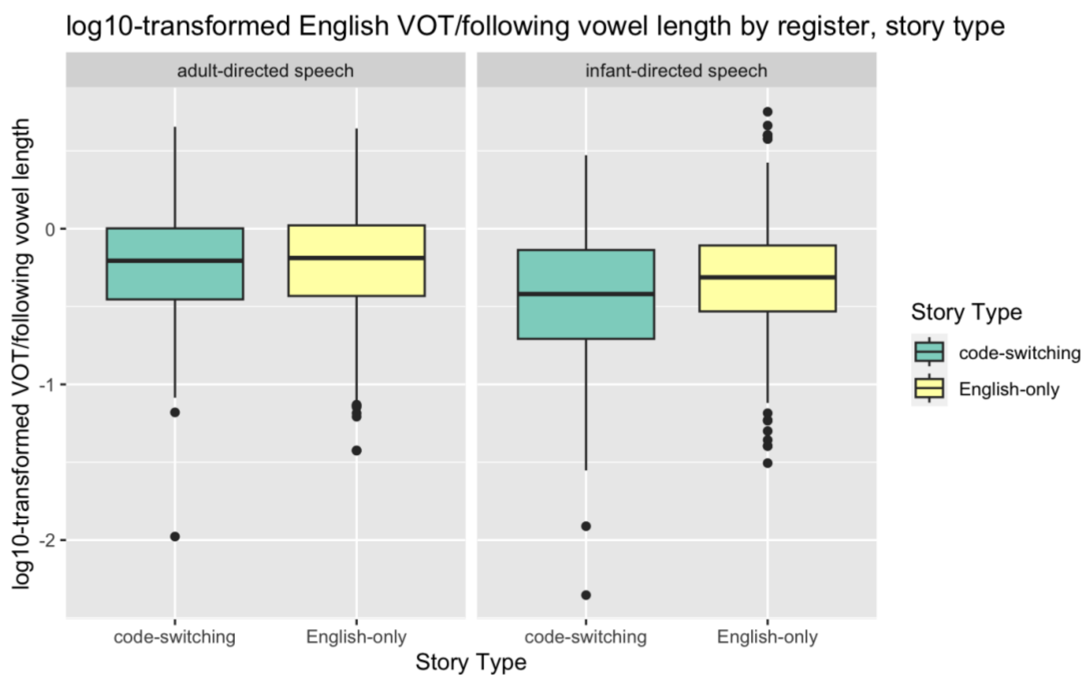


Figure 18: Production study - Analysis 2: English consonant-vowel ratio by register, story type – log10 transformed



For log-transformed VOT as the dependent variable, this analysis suggested a main effect of word length, such that longer words were produced with longer VOTs ($t(1923) = 10.795$, $\beta = 0.231$, $p < 0.001$. This is likely an effect of speaking rate, since slower speech would be expected to lead to both longer words and longer VOTs. However, as word length is only measured using duration here, rather than syllable count, it is possible that these effects of word length are the result of both speaking rate and because some words are multi-syllabic. There was also a significant interaction between register and story type ($t(1710) = 1.982$, $\beta = 0.0252$, $p < 0.05$. This interaction is further assessed below. There was no significant main effect of register ($p = 0.212$) or story type ($p = 0.949$).

For log-transformed consonant-vowel ratio as the dependent variable, there was again a main effect of word length, though it was in the opposite direction. That is, after normalizing for speech rate by dividing VOT by the length of the following vowel, longer words were produced with *shorter* VOTs ($t(1896) = -4.797$, $\beta = -0.147$, $p < 0.001$). This analysis also suggested a main effect of register, such that adult-directed speech had longer VOTs than infant-directed speech, after normalizing for speech rate ($t(6.79) = 3.206$, $\beta = 0.116$, $p < 0.05$, as well as a marginally significant interaction between register and story type ($t(1214) = 1.898$, $\beta = 0.034$, $p = 0.058$).

The interaction between register and story type was assessed using a post-hoc exploratory analysis. Although splitting the data based on register and then assessing the effect of story type for each register would be a better assessment of the current research question, doing so did not yield results that help to interpret this interaction. Therefore, the effect of register is tested separately for the English-only and the code-switching stories. Four linear mixed-effects models were fit: English-only stories with VOT as dependent variable, English-only stories with consonant-vowel ratio as dependent variable, code-switching stories with VOT as dependent variable, code-switching stories with consonant-vowel ratio as dependent variable. The maximal random effects structure that converged for all four models included random intercepts of both phoneme and participant and random slopes of register for each participant.

Code-switching stories

The analysis of code-switching stories did not show a significant effect of register using either VOT ($t(3.58) = 1.88$, $\beta = 0.082$, $p = 0.142$) or consonant-vowel ratio ($t(7.47) = 1.735$, $\beta = 0.108$, $p = 0.124$ as the dependent variable. In both cases, there is a significant effect of word length in the same direction as the full model described above (positive when using VOT as the dependent variable, negative when accounting for speech rate by dividing VOT by the length of the following vowel).

English-only stories

The analysis of English-only stories using VOT as the dependent variable also did not show an effect of register ($t(9.033) = 0.363$, $\beta = 0.0174$, $p = 0.725$. However, there was a marginally significant effect of register when using consonant-vowel ratio as the dependent variable, $t(9.05) = 2.08$, $\beta = 0.08$, $p = 0.067$, such that adult-directed speech had longer VOTs than infant-directed speech after accounting for speech rate. Again, both models showed significant effects of word length in the same direction as the analyses described above. This post-hoc analysis therefore showed tentative evidence that the interaction between register and story type is driven by longer voice onset times in adult-directed speech particularly in the English-only stories[13]. However, because these effects were not statistically significant at an alpha-level of 0.05 and results are preliminary with only 10 participants, this pattern may not be robust and should not be meaningfully interpreted.

[13] Mathematically, it is also possible that this effect resulted from similar VOT lengths in ADS and IDS, but a faster speech rate (with shorter vowels) in ADS, which thus led to higher consonant-vowel ratios for ADS, though this possibility is unlikely given the relationship between speech rate and VOT length.

Analysis 3: Moderately local effect of code-switching on VOT (utterance type; unilingual vs. code-switched)

The analysis of the moderately local effect of code-switching included all tokens of English voiceless stops produced in the code-switching stories, and compared VOTs of tokens produced in unilingual utterances to those produced in code-switched utterances. A total of 665 tokens were included in this analysis, across all 10 participants and 3 phonemes: 310 that were removed from Spanish by at least 3 seconds ("unilingual utterances") and 355 that were produced within 3 seconds of Spanish ("code-switched utterances"). These tokens were fairly split between adult-directed and infant-directed speech, though slightly more unilingual tokens were produced in adult-directed speech than infant-directed speech (175 vs. 135 tokens).

Figure 19: Production study - Analysis 3: English VOT (ms) in code-switched stories by register, utterance type

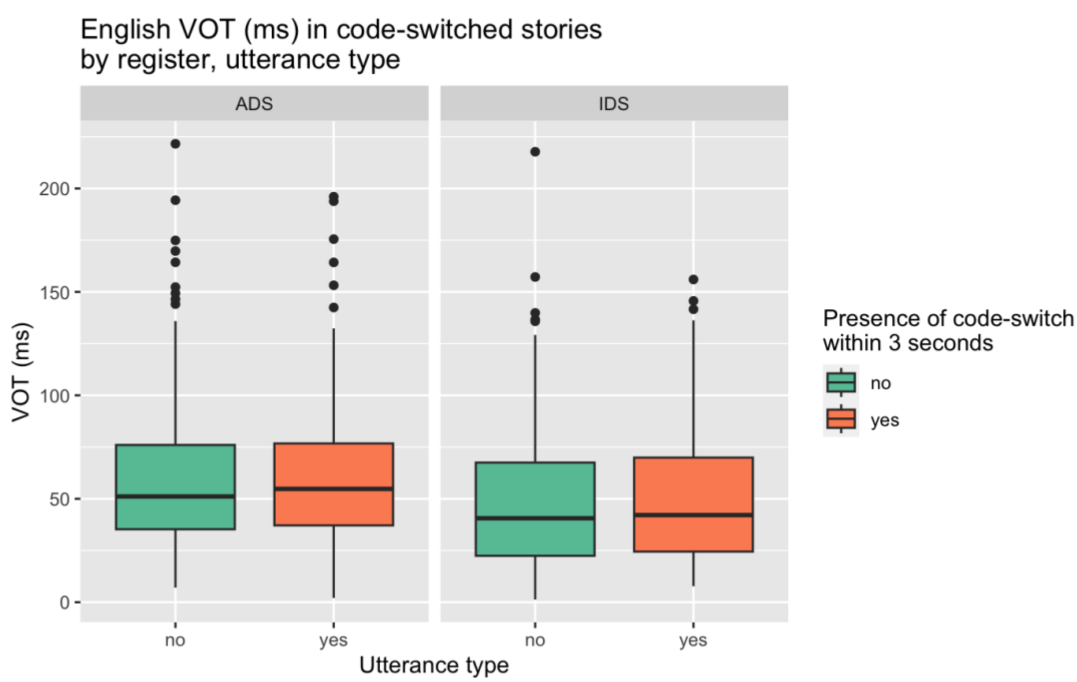


Figure 20: Production study - Analysis 3: English VOT (ms) in code-switched stories by register, utterance type – log10-transformed

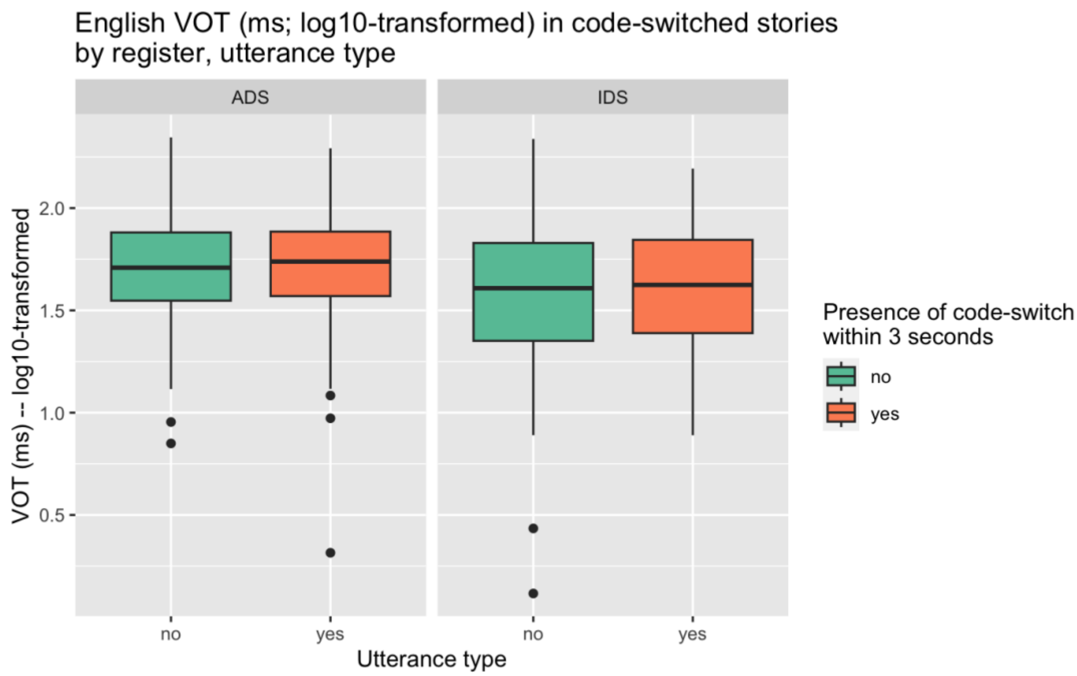


Figure 21: Production study - Analysis 3: English consonant-vowel ratio in code-switched stories by register, utterance type

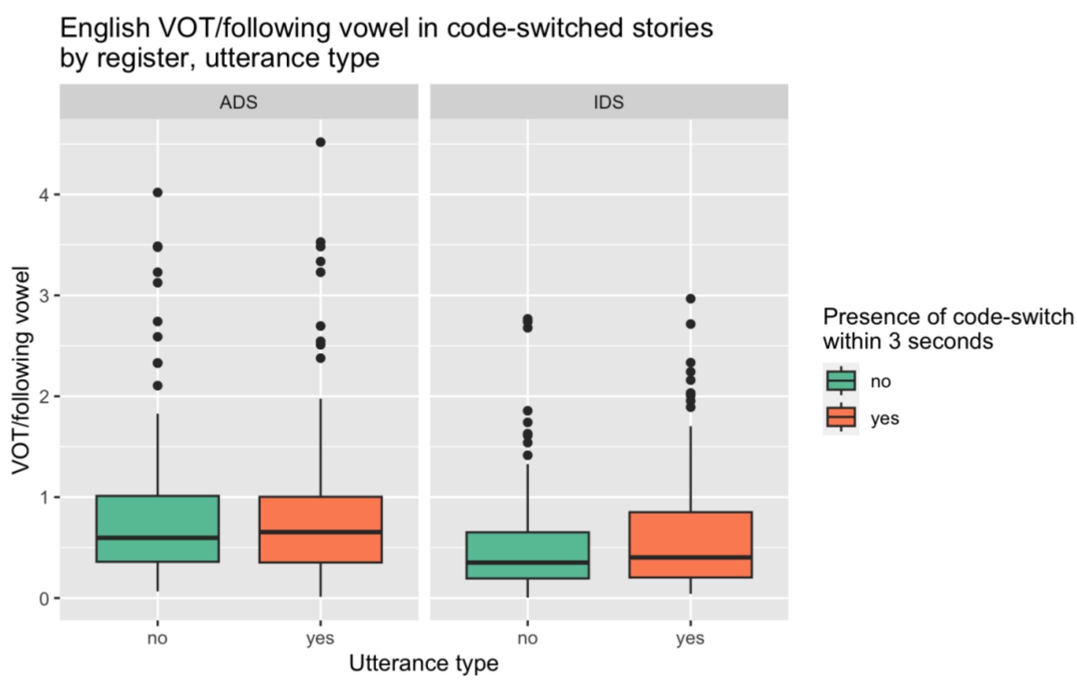
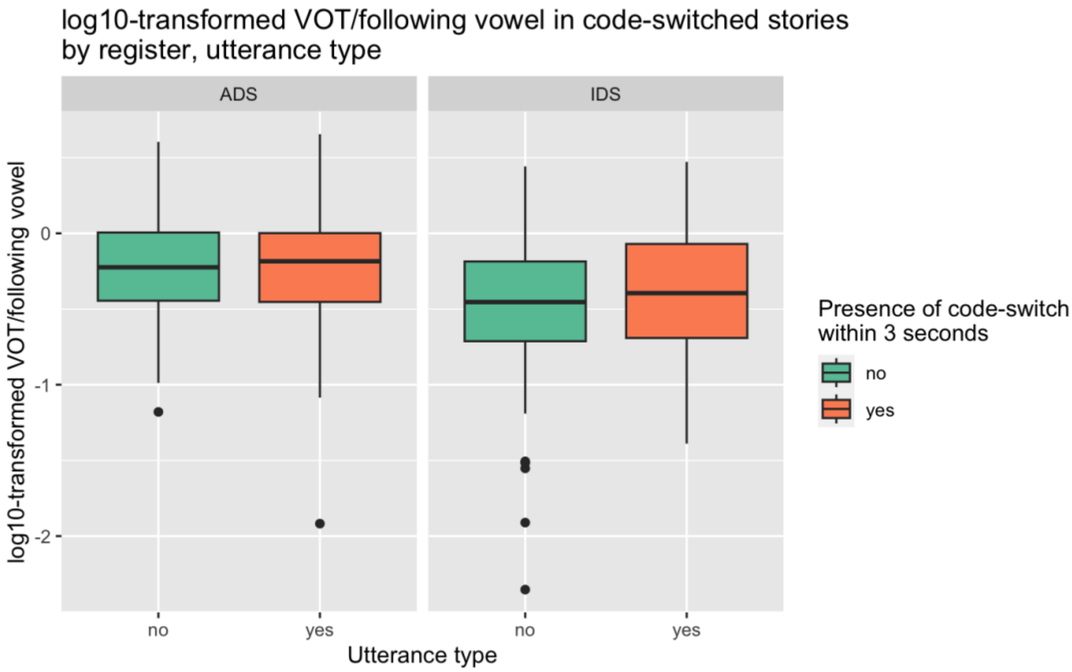


Figure 22: Production study - Analysis 3: English consonant-vowel ratio in code-switched stories by register, utterance type – log10-transformed



The effects of register (infant-directed vs. adult-directed), utterance type (unilingual vs code-switched) and the interaction between these two factors were assessed while controlling for word length. Register and utterance type were effects-coded (-1 and 1). Word length was again expressed in terms of milliseconds and log-transformed. The maximal random effects structure which converged for the analyses of VOT and consonant-vowel ratio included random intercepts of both phoneme and participant, and random slopes of register for each participant.

Analyses using both VOT and consonant-vowel ratio as dependent variables showed only a significant effect of word length, but no effect of register or utterance type, or an interaction between word length and either of these primary effects of interest. Both effects of word length were in the same direction as previously reported, such that in the analysis of VOT, longer words had longer voice onset times ($t(644.5) = 5.888$, $\beta = 0.247$, $p < 0.001$) and in the analysis of consonant-vowel ratio, longer words had shorter voice onset times ($t(651.9) = -2.967$, $\beta = -0.177$, $p < 0.01$).

Analysis 4: Local effect of code-switching on VOT (distance to and direction of switch)

The analysis of the local effect of code-switching included only tokens of English voiceless stops from the code-switching stories which were produced within 3 seconds of Spanish. This analysis includes a total of 355 tokens. For 193 of these tokens, the closest Spanish to the token was produced after the token, such that any effects of Spanish on that voice onset time would be anticipatory in nature. For the remaining 162 tokens, the closest Spanish was produced before the token such that any effects of Spanish would be perseverative in nature. However, for 24 of these "anticipatory" tokens and 42 of these "perseverative" tokens, Spanish was spoken within three seconds of the token in both directions (both preceding and following it). These 66 potentially ambiguous tokens are only included in the category to which they best belong, though Fricke et al. (2016) included similarly ambiguous tokens in their analyses of both anticipatory and perseverative effects.

The effects of register (infant-directed vs. adult-directed), switch direction (English-to-Spanish, such that effects of Spanish on the English token are anticipatory vs. Spanish-to-English, such that effects of Spanish on the English token are perseverative), distance to the closest instance of Spanish (in seconds) and the interaction between these three factors were assessed while controlling for word length. Register and switch direction were effects-coded (-1 and 1). Word length was again expressed in terms of milliseconds and log-transformed. Distance to the nearest code-switch was expressed in seconds. This distance was not transformed, as it approximately followed a normal distribution. The maximal random effects structure which converged for the analyses of VOT and consonant-vowel ratio included random intercepts of both phoneme and participant, and random slopes of register for each participant.

Using VOT as the dependent variable, the only significant effect was the effect of word length, such that longer words had longer voice onset times ($t(335.3) = 2.966$, $\beta = 0.17$, $p <$

0.05). When controlling for speech rate by dividing by the length of the following vowel, though, the model suggested a significant effect of switch direction such that tokens that preceded Spanish and thus could have anticipatory effects of Spanish were produced with a shorter voice onset time than tokens that followed Spanish and thus could have perseverative effects ($t(330.4) = -2.116$, $\beta = -0.172$, $p < 0.05$), as well as a marginally significant interaction between distance and switch direction ($t(331.2) = 1.756$, $\beta = 0.092$, $p = 0.08$).

This potential interaction was assessed by splitting the data into four groups: adult-directed tokens that precede Spanish, adult-directed tokens that follow Spanish, infant-directed tokens that precede Spanish, infant-directed tokens that follow Spanish. In all four cases, I tested for an effect of distance to the nearest Spanish on the length of the voice onset time, while accounting for the length of the word. The data was split based on both register and switch direction, despite a lack of a significant effect of register in the initial models, in order to allow for a straightforward comparison between the present data and previous studies which have only considered adult-directed speech. The first two analyses are therefore a conceptual replication of Fricke et al. (2016), which considered the two switch directions separately and found an effect of distance in the English-to-Spanish direction but not the Spanish-to-English direction. The second two analyses test whether similar effects are found in infant-directed speech.

The only random effects structure that converged for all eight models (both dependent variables for each of the four groups of data) included random intercepts of both phoneme and participant, but no random slopes.

<u>Adult-directed, precede Spanish (anticipatory effects of Spanish)</u>

91 tokens were produced in adult-directed speech and preceded Spanish by less than three seconds. In both the analysis of VOT and of consonant-vowel ratio, there is a significant effect of distance such that tokens that were produced further from Spanish had a longer VOT,

and tokens that were produced closer to Spanish had a shorter VOT (VOT as dependent variable: $t(77.82) = 2.249$, $\beta = 0.144$, $p < 0.05$; consonant-vowel ratio as dependent variable: $t(78.88) = 2.092$, $\beta = 0.203$, $p < 0.05$). There was also a significant effect of word length for the analysis of VOT, such that longer words had longer VOTs ($t(79.39) = 2.011$, $\beta = 0.19$, $p < 0.05$), but not for the analysis of consonant-vowel ratio.

For a word of average length in this subset of data (434.5 ms), a token produced 0 seconds preceding Spanish has an expected VOT of 38.4 ms and consonant-vowel ratio of .351 according to this model. A token produced in a word of the same length but which precedes Spanish by 3 seconds has an expected VOT of 59.1 ms and a consonant-vowel ratio of .646. The average vowel length for this set of tokens is 110 ms, such that these consonant vowel-ratios correspond to VOTs of 38.7 and 71.1 ms, respectively. Note that this range of VOTs from 38.4 to 71.1 ms is entirely within the standard range of VOTs for English voiceless stops, but 38.4 and 38.7 are closer to the range associated with Spanish voiceless stops (~0-30 ms) than 59.1 and 71.1 are.

<u>Infant-directed, precede Spanish (anticipatory effects of Spanish)</u>

102 tokens were produced in infant-directed speech and preceded Spanish by less than three seconds. There was not a significant effect of distance in either the analysis of VOT ($p = 0.998$) or consonant-vowel ratio ($p = 0.662$). Numerically, both relationships were negative such that tokens produced closer to Spanish had longer VOTs, though this was not significant. The analysis of consonant-vowel ratio did show a significant effect of word length, such that when accounting for speech rate by dividing VOT by the length of the following vowel, longer words had shorter consonant-vowel ratios ($t(93.65) = -3.003$, $\beta = -0.512$, $p < 0.01$). There was no effect of word length for the analysis of VOT.

Figure 23: Production study - Analysis 4: English VOTs in code-switched utterances preceding Spanish

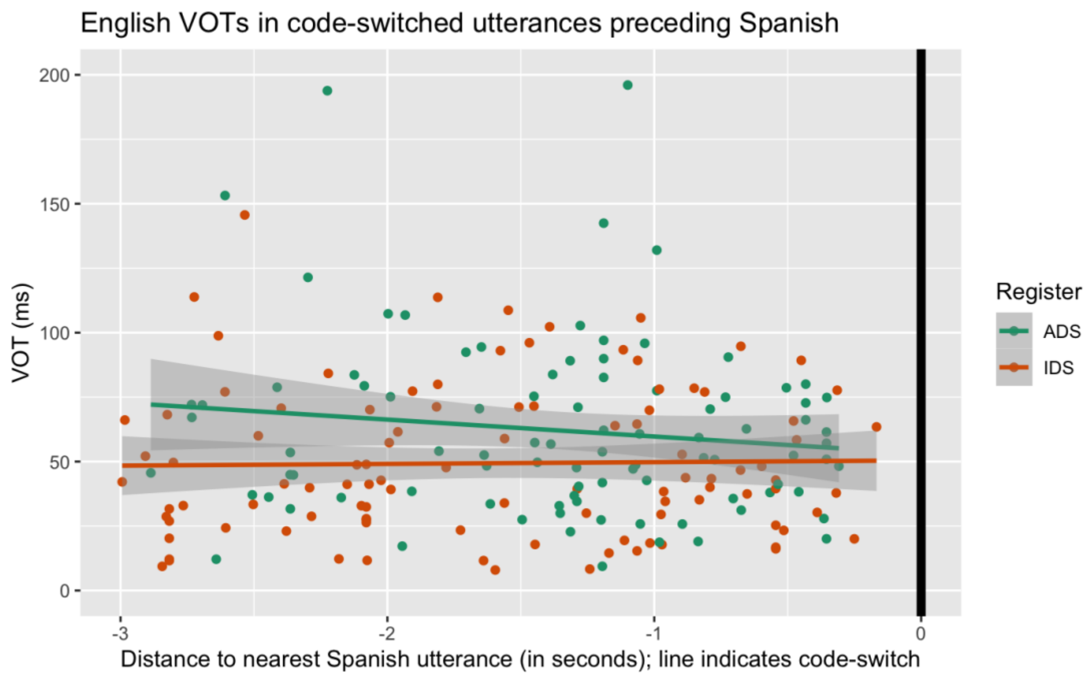


Figure 24: Production study - Analysis 4: English VOTs in code-switched utterances preceding Spanish (log10 transformed)

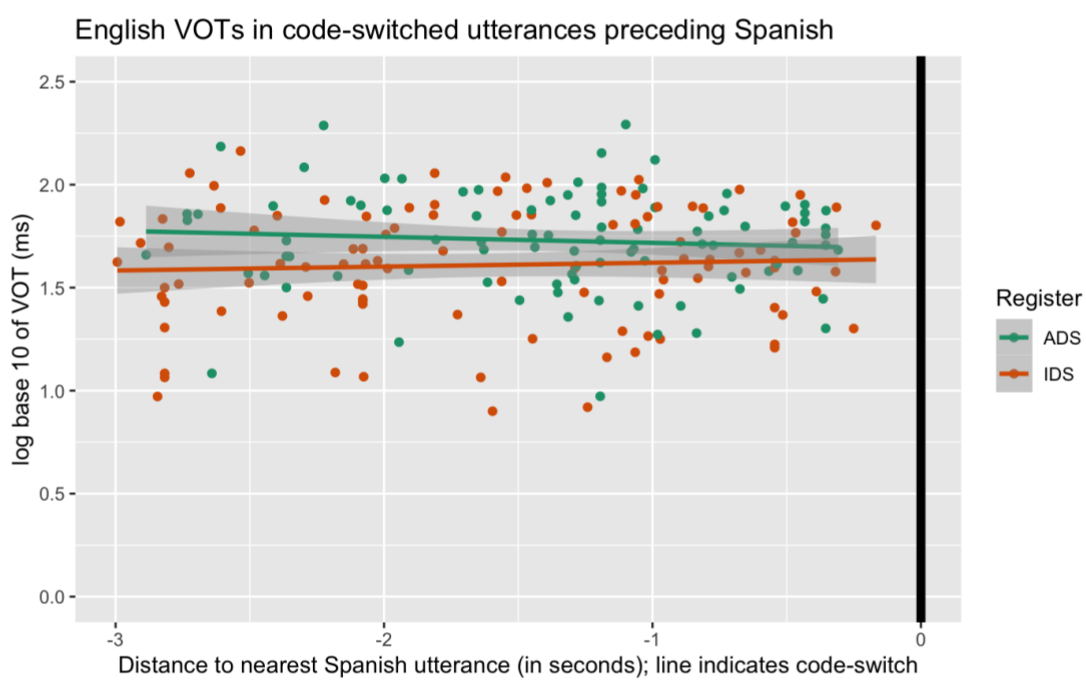


Figure 25: Production study - Analysis 4: English consonant-vowel ratio in code-switched utterances preceding Spanish

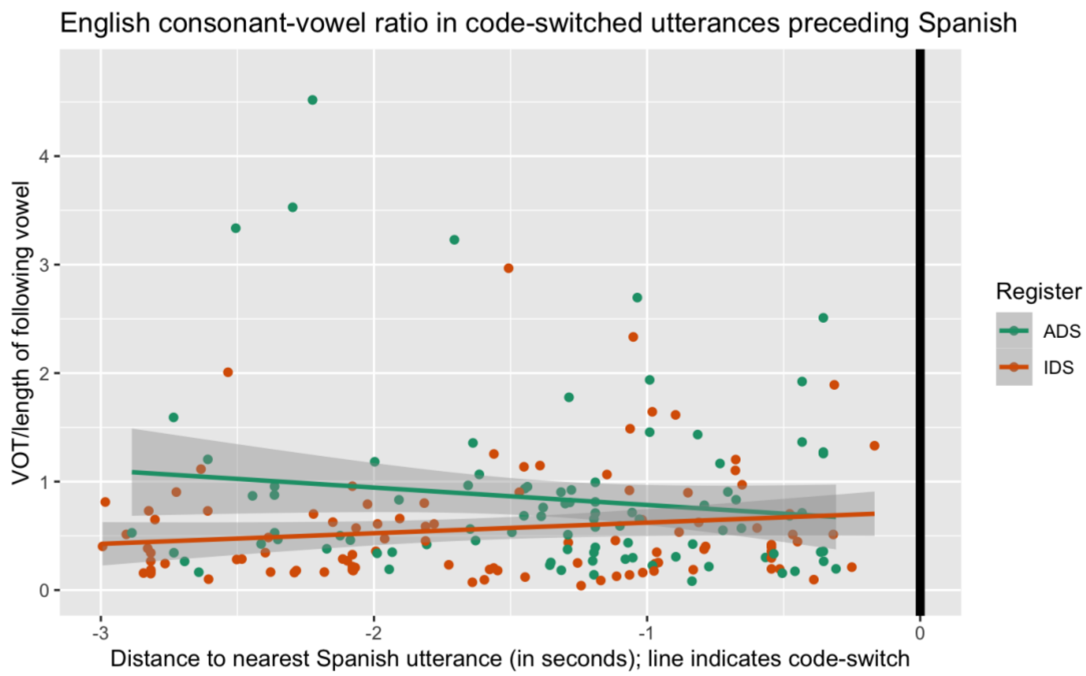
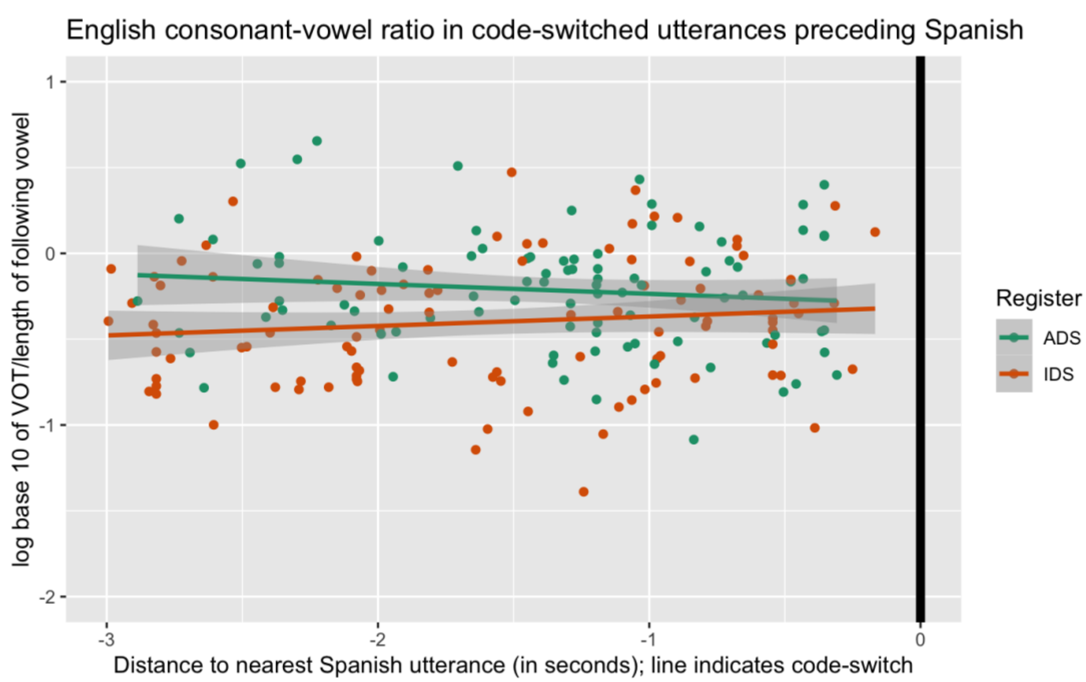


Figure 26: Production study - Analysis 4: English consonant-vowel ration in code-switched utterances preceding Spanish (log10 transformed)



Adult-directed, follow Spanish (perseverative effects of Spanish)

87 tokens were produced in adult-directed speech and followed Spanish by less than three seconds. There was not a significant effect of distance in either the analysis of VOT ($p = 0.672$) or consonant-vowel ratio ($p = 0.222$). Numerically, both relationships were negative such that tokens produced closer to Spanish had longer VOTs, though this effect was not significant. The analysis of VOT did show a significant effect of word length, such that longer words had longer VOTs ($t(78.88) = 2.677$, $\beta = 0.32$, $p < 0.01$. The combined findings from these two analyses of adult-directed speech serve as a conceptual replication of Fricke et al. (2016), which will be further addressed in the discussion section.

Infant-directed, follow Spanish (perseverative effects of Spanish)

75 tokens were produced in infant-directed speech and followed Spanish by less than three seconds. There was not a significant effect of distance in either the effect of VOT ($p = 0.524$) or consonant-vowel ratio ($p = 0.251$). Numerically, both relationships were negative such that tokens produced closer to Spanish had longer VOTs, though this was not significant. There was also not a significant effect of word length at an alpha-level of 0.05 in either analysis, though the effect of word length was marginally significant when using VOT as the dependent variable ($t(68.24) = 1.746$, $\beta = 0.188$, $p = 0.085$), such that longer words may have longer VOTs.

Figure 27: Production study - Analysis 4: English VOTs in code-switched utterances following Spanish

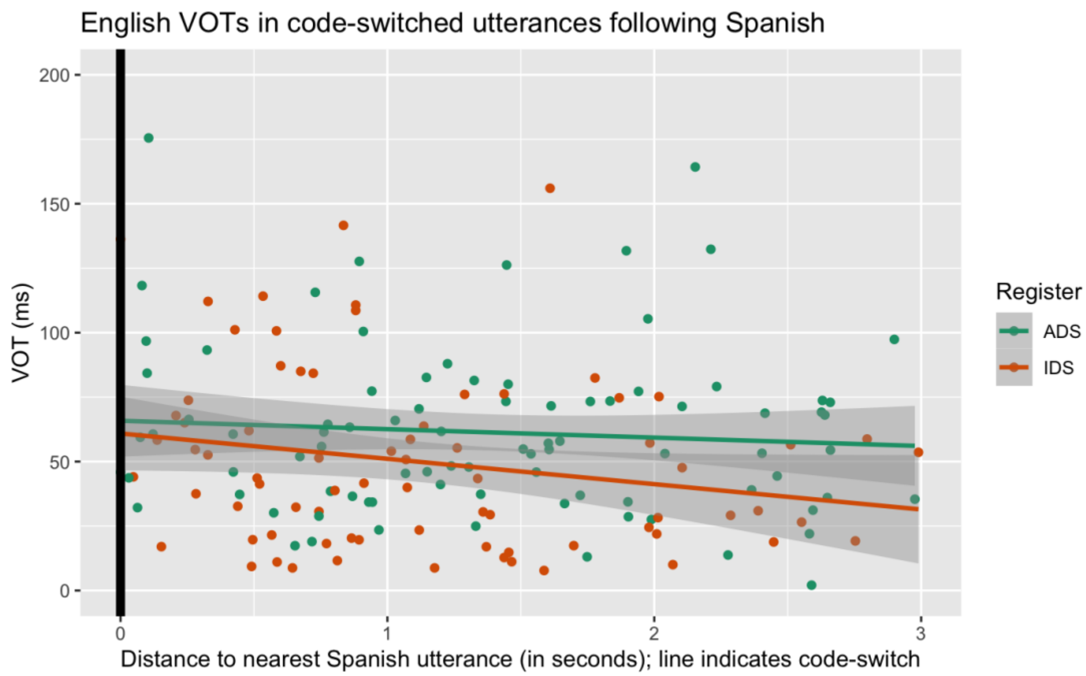


Figure 28: Production study - Analysis 4: English VOTs in code-switched utterances following Spanish (log10 transformed)

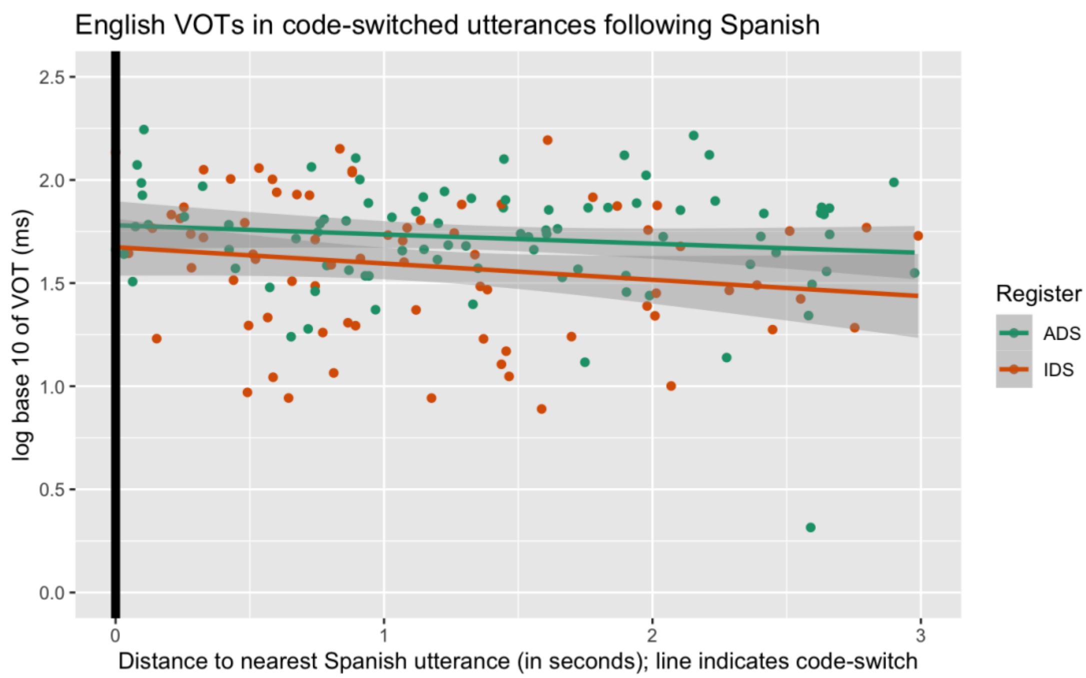


Figure 29: Production study - Analysis 4: English consonant-vowel ratio in code-switched utterances following Spanish

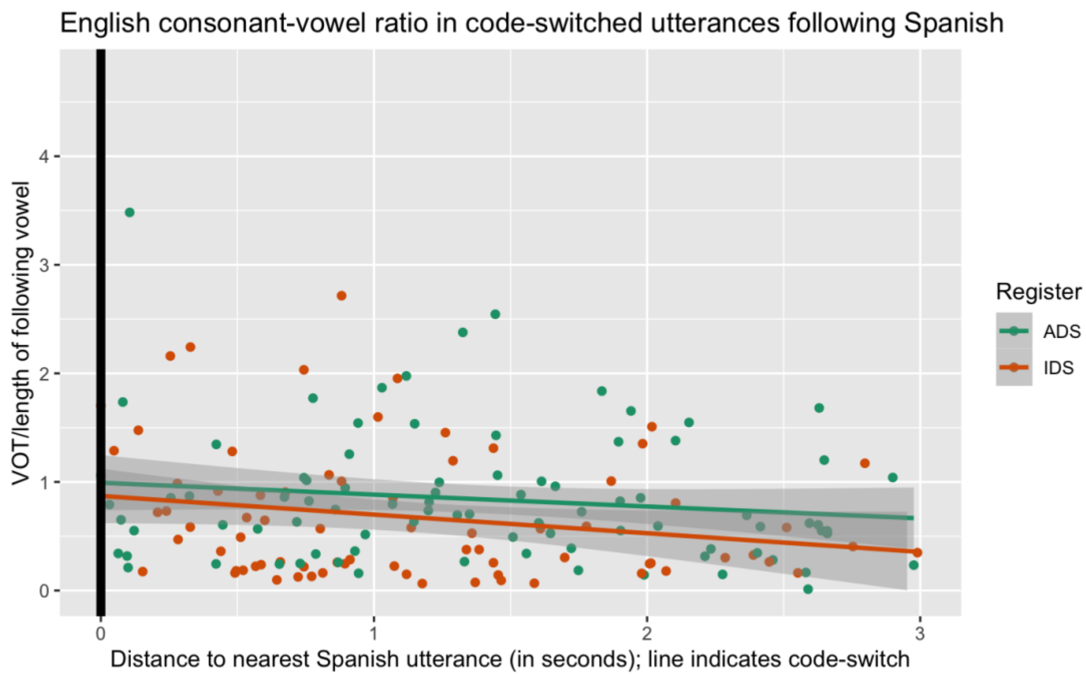
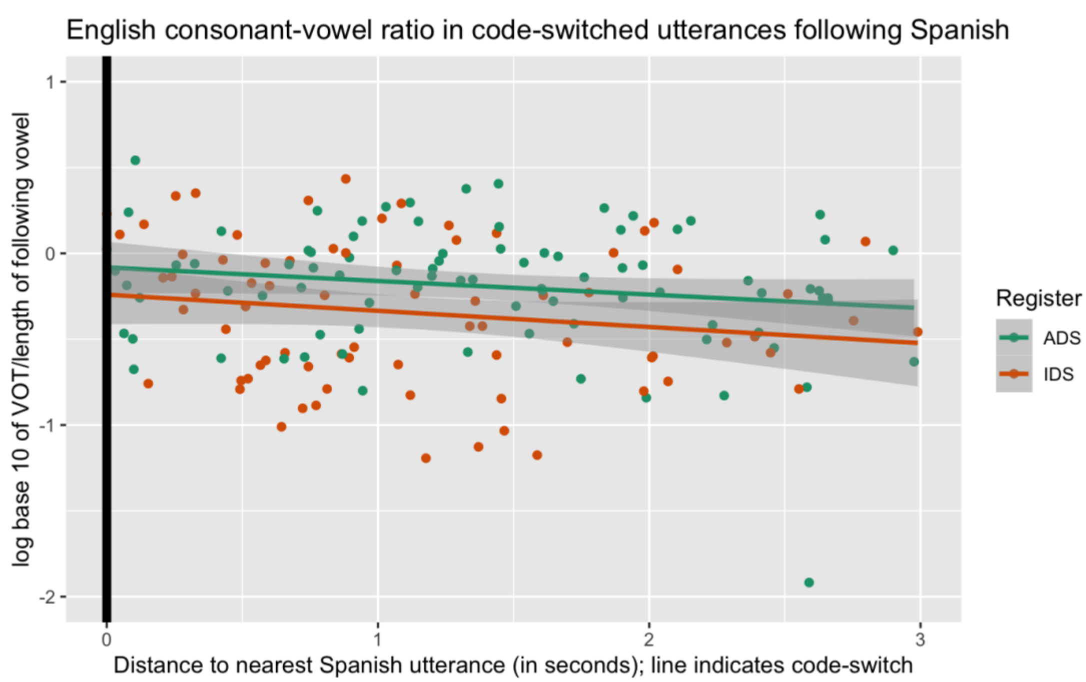


Figure 30: Production study - Analysis 4: English consonant-vowel ratio in code-switched utterances following Spanish (log10 transformed)



Analysis 4b: Local effect of code-switching, but expanding the range of analysis to 5 seconds

In analysis 4 as described above, the dataset is limited to English tokens produced within 3 seconds of Spanish, and tokens that precede and follow Spanish were compared. This analysis showed some evidence of an anticipatory effect of Spanish on English VOTs in adult-directed speech, such that when speakers were talking to an adult and about to switch into Spanish, their English VOTs were shorter and more Spanish-like. In analysis 4b, I expanded the range to 5 seconds to test whether the observed effect is an unintended artifact driven by the initially selected range of 3 seconds. In both Fricke et al (2016) and Balukas & Koops (2015), the majority of tokens occurred within 5 seconds of a code-switch, including many from 3-5 seconds, which provides supporting evidence for the selection of 5 seconds. Moreover, given that IDS is typically slower than ADS, it might make sense to use a wider range when considering effects in the infant-directed stimuli. By expanding the range in this secondary analysis we are also able to further assess the time course of the effect of code-switching.

Expanding the range of analysis to five seconds leads to a total of 266 tokens which precede Spanish. 63 of these are potentially ambiguous in that they also follow Spanish by less than 5 seconds, though the token-to-Spanish distance is shorter than the Spanish-to-token distance for the tokens included in this “anticipatory” analysis. Expanding the range to 5 seconds also leads to a total of 226 tokens which follow Spanish, 78 of which are potentially ambiguous.

The same patterns as observed in analysis 4a were replicated in analysis 4b, using a longer range of analysis, though perhaps to a weaker extent. Across all 492 tokens in the analysis of consonant-vowel ratio, there was a significant main effect of switch direction such that tokens followed by Spanish have shorter VOTs accounting for speech rate than those preceded by Spanish ($t(468.81) = -2.421$, $\beta = -0.146$, $p < 0.05$), and a significant interaction between distance and switch direction ($t(469.12) = 2.149$, $\beta = 0.053$, $p < 0.05$. These effects did not appear in the analysis of VOT (p-values = 0.421 and 0.533, respectively), though they are numerically in the same directions. Both the analyses of VOT and consonant-vowel ratio show the effect of word length that has been described in all previous analyses.

The critical sub-analysis for replication purposes is that of adult-directed speech preceding Spanish. 135 tokens were produced in adult-directed speech and preceded Spanish by less than five seconds. In the analysis of consonant-vowel ratio, there is a significant effect of distance such that tokens that were produced further from Spanish had a longer VOT, and tokens that were produced closer to Spanish had a shorter VOT ($t(123.37) = 2.357$, $\beta = 0.095$, $p < 0.05$). However, this effect was not significant in the analysis of VOT that does not account for speech rate ($t(122.67) = 1.292$, $\beta = 0.037$, $p = 0.199$. Both analyses show the previously-described effect of word length. According to this model, the consonant-vowel ratio associated with a token from a word of average length which is produced 0 seconds preceding Spanish is .398, which corresponds to a VOT of 42 ms if the vowel is of average length. For a token with the same word length and vowel length which is produced 5 seconds preceding Spanish, this model estimates a consonant-vowel ratio of .641, or a VOT of 67.6 ms.

Figure 31: Production study - Analysis 4b: VOT in code-switched adult-directed speech preceding Spanish using a 5-second analysis window

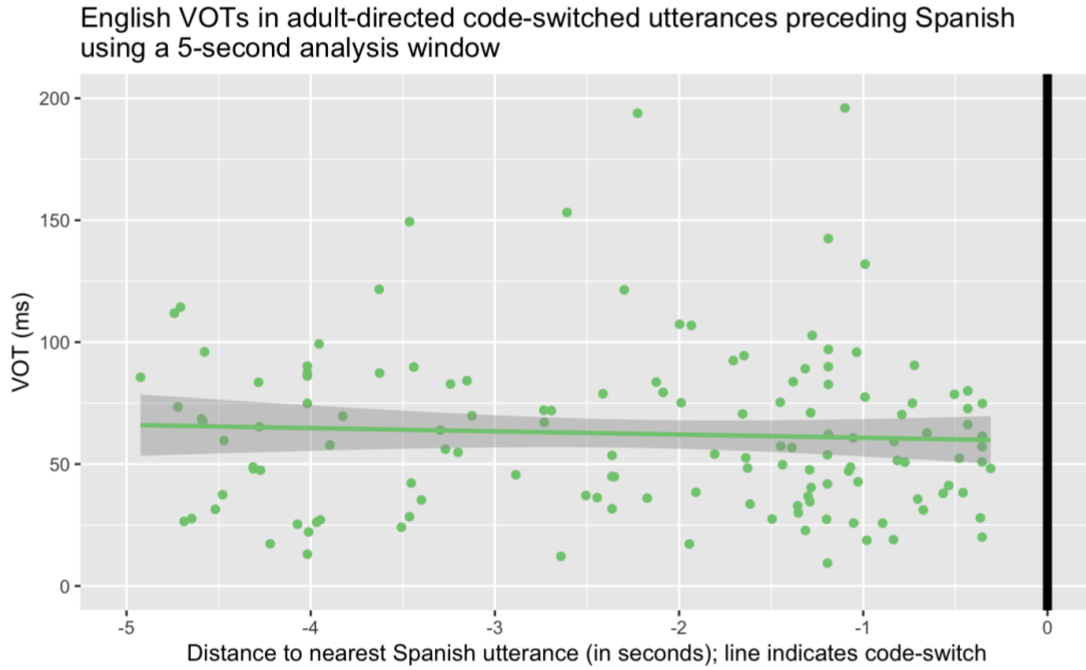


Figure 32: Production study - Analysis 4b: log10-transformed VOT in code-switched adult-directed speech preceding Spanish using a 5-second analysis window

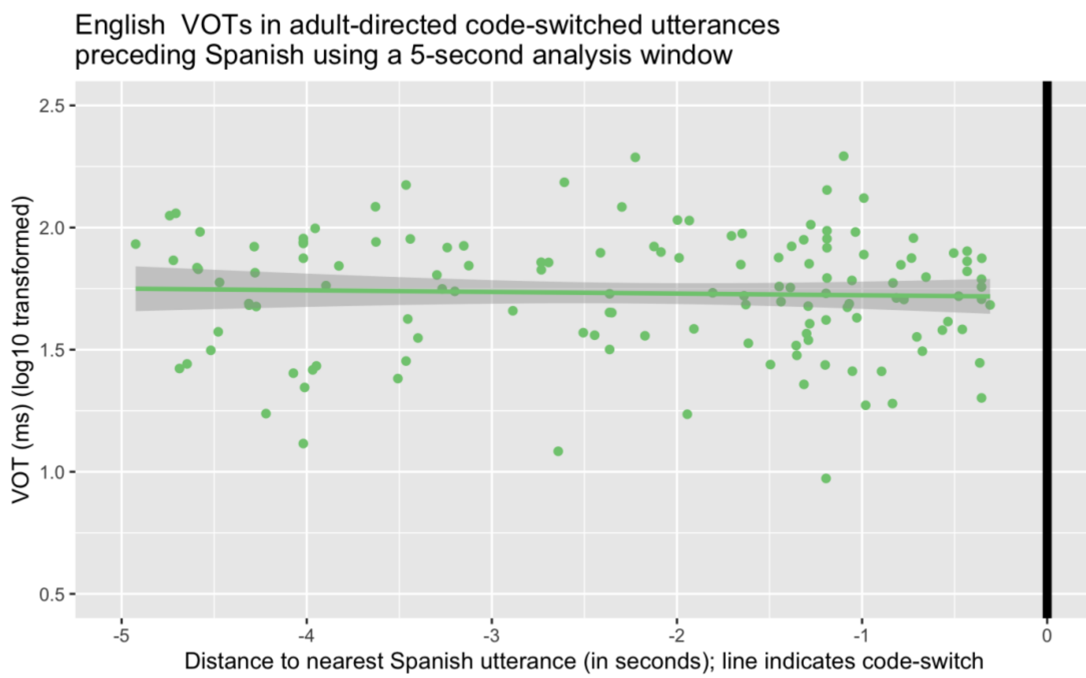


Figure 33: Production study - Analysis 4b: VOT/following vowel in code-switched adult-directed speech preceding Spanish using a 5-second analysis window

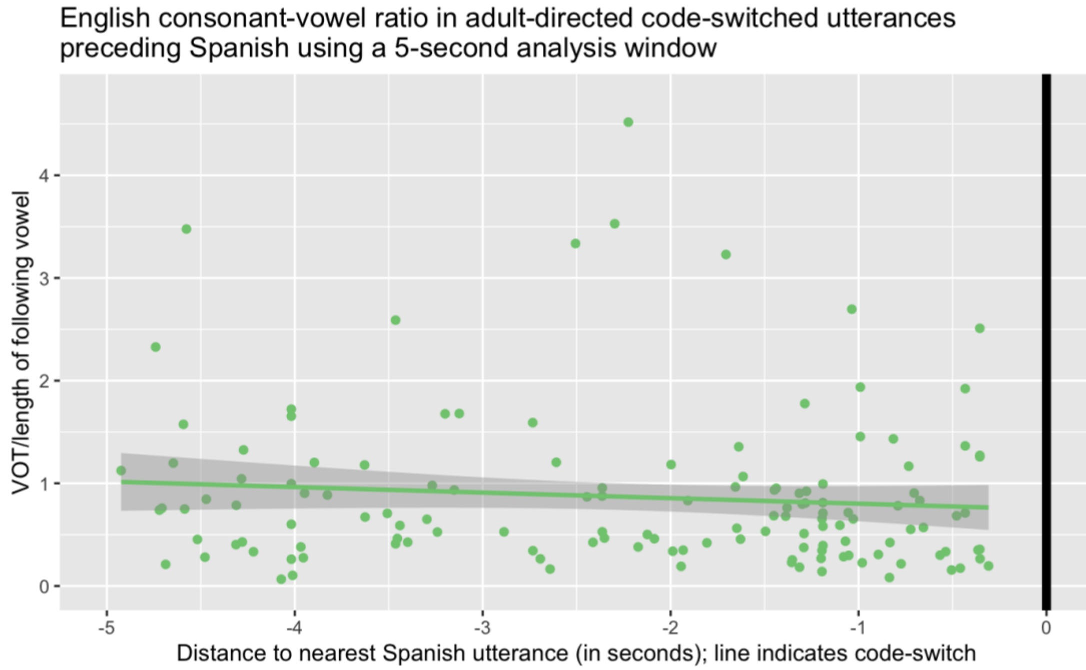
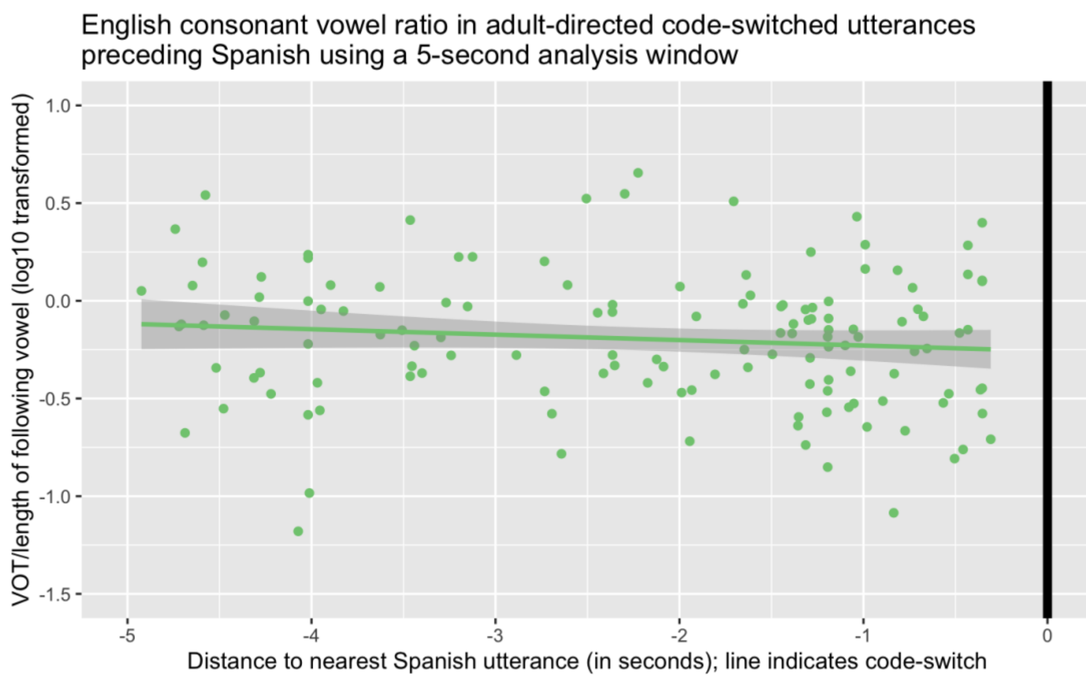


Figure 34: Production study - Analysis 4b: log10-transformed VOT/following vowel in code-switched adult-directed speech preceding Spanish using a 5-second analysis window



In all other analyses there was not a significant effect of distance on either VOT or consonant-vowel ratio. Numerically all relationships are negative, such that if anything, English tokens produced closer to a code-switch may have slightly longer VOTs. Analysis of additional data to reach the target sample size of 20 participants will be necessary to fully assess this relationship between distance and VOT in infant-directed speech and in adult-directed speech that follows a code-switch. Overall, the pattern observed in analysis 4b suggests that the effect observed at a three-second range may extend to five seconds, but potentially to a weaker degree.

Discussion

In this experiment, ten Spanish-English bilingual women each narrated six wordless picture books at home in a natural, comfortable environment. Three of these stories were told to their infant, ranging from 6 to 20 months of age. The other three were told to a familiar Spanish-English bilingual adult. To each person, participants told one story using only English, one using only Spanish, and one while freely switching between English and Spanish. I focused on when participants were speaking English and analyzed their average fundamental frequency, and the distribution of voice onset times in each language context. Data collection and additional analysis is ongoing in order to reach a full sample of 20 participants and to analyze both English and Spanish productions; however, prior research suggests that English VOTs may be more likely to differ due to language context than Spanish VOTs.

In analysis 1, I replicated a classic characteristic of infant-directed speech and found that fundamental frequency was both higher and more variable in infant-directed speech as compared to adult-directed speech. In analyses 2-4, analyses move from a relatively global language context (language mode; English-only vs. code-switching) to a more local language context (within three seconds of a code-switch) and test effects of language and sentence context on the voice onset times of English voiceless stops. This work provides no evidence that the production of English voiceless stops is affected by the relative mental activation of Spanish in infant-directed speech, but some tentative evidence that the production of English voiceless stops is affected by the relative mental activation of Spanish in adult-directed speech. This work also replicated evidence for effects of speech rate and word length on voice onset time. In the following sections I expand on each of these points in turn.

Anticipatory and perseverative effects of code-switching

In these analyses I found some evidence for anticipatory effects of code-switching on the production of English voiceless stops in adult-directed speech. Specifically, when we focus on tokens of English voiceless stops in adult-directed speech that are produced up to three seconds before a code-switch into Spanish, there is an effect of distance such that tokens produced close to a switch have a shorter (more Spanish-like) voice onset time than those produced farther from a switch. This effect holds, but may be slightly weaker, when we expand the range of analysis to five seconds. In the opposite direction, switches from Spanish into English, there is no evidence of a significant effect of distance to the code-switch on English VOTs. If anything, English voiceless stops produced after a Spanish-to-English code-switch may have had *longer* voice onset times in IDS when they are closer to that code-switch than when they were more temporally separated from Spanish, though this effect does not approach significance with 10 participants.

The results here in adult-directed speech are in line with those discussed in Fricke, et al. (2016), who likewise found an effect of distance on the VOT of tokens of English voiceless stops produced before a code-switch but not after a code-switch. Analysis 4 accounted for speech rate differently than Fricke and colleagues, by dividing the VOT by the length of the vowel (or approximant) which follows it, rather than calculating articulation rate over the utterance, but find concurrent results. These results therefore corroborate their findings and establish that the observed effect is unlikely to be an artifact driven by minor methodological choices.

When considering the adult-directed speech alone, a reasonable conclusion (and indeed, the conclusion discussed in Fricke et al. (2016) is that such anticipation of a code-switch is a natural by-product of the preparation process: “Taken together, a possible interpretation of our findings is that of active anticipation and suppression on the part of the speaker; a speaker who intends to switch languages may begin preemptively increasing the activation of the language not in use (giving rise to anticipatory cross-language phonological influence) while simultaneously suppressing the activation of the language currently in use (yielding slightly slowed speech rate), allowing them to seamlessly switch into the other language (and show no signs of perseverative activation of the previously active language)” (Fricke et al., 2016, p. 120).

This conclusion is further supported by a story re-telling task in French and English reported by Grosjean & Miller (1994). When participants were told to re-tell a story using French, but to produce character names in English, the word-initial voiceless VOTs of those character names did not significantly differ than when the story was told entirely in English. Under the terminology used in the present study, that analysis was essentially testing for perseverative effects of French on English at the point of a single-word insertion, and such effects were not found. As participants knew that they were expected to use English pronunciation when producing character names, and could decide exactly when to use those character names, they may have been able to prepare for English by increasing its level of activation before the point of the code-switch. Grosjean & Miller (1994) did not directly test for anticipatory effects of code-switching, or allow participants to use any code-switches that were not single-word insertions of character names, so the broader applicability of that study to the present question is somewhat limited.

Here I provided additional evidence that when bilinguals freely code-switch, engaging in a natural language behavior, activation of one language may occur at the phonological level before an observable "switch" at the lexical or syntactic level. That is, if a bilingual is about to switch from English into Spanish, they may essentially ramp up activation of Spanish representations before producing a lexical item and syntactic structure in Spanish, so that the switch itself can be fluent. Activation of Spanish does not, however, appear to persist after a switch from Spanish into English.

The next steps for this research involve analyzing production in Spanish in order to assess whether the same effect of Spanish activation on English productions occurs in the other direction, leading to an effect of English on Spanish. Patterns in Spanish which mirror those observed in English would strengthen this argument. If when a bilingual is speaking in Spanish but about to switch into English, activation of English representations increases and Spanish productions become more English-like, this pre-existing phonological activation of English would result in the observed lack of Spanish perseveration following a code-switch. Likewise, a lack of English perseveration on Spanish productions following a code-switch could be explained by the increasing activation of Spanish *before* a lexical switch occurs.

In infant-directed speech, however, the patterns differed. In this speech register, there was no evidence that voice onset times of English voiceless stops were shorter immediately preceding a code-switch into Spanish than when they preceded a code-switch by 3 or 5 seconds. Numerically the effect is actually such that the VOTs of English voiceless stops *increase* and become more English-extreme as they approach a code-switch into Spanish. Though this effect does not approach significance, and therefore should not be meaningfully interpreted with a sample size of 10 participants, it is worth discussing because of the predictions it makes about the expected results with a full sample. That is, the effect of distance to a code-switch in infant-directed speech is unlikely to be so strong in the 10 remaining-to-be-collected participants but in the opposite direction (showing that English VOTs are shorter when closer to a code-switch) such that it would flip the direction of the current effect and match the pattern currently observed in adult-directed speech. It is more likely that either the observed effect will remain in the current direction and that doubling the sample size will increase power and lead to a significant effect, or that the results with a full 20 participants will remain somewhat variable and result in a null effect. Likewise in the adult directed speech, it is likely that 10 additional participants will simply serve to strengthen the current observed effect. While it is possible that they could instead be more variable and lead to a null effect, it is very unlikely that the current significant effect in adult-directed speech that immediately precedes a code-switch would flip directions and reach significance in the opposite direction. Therefore, we expect that there will be a difference in the patterns between infant-directed and adult-directed speech such that there is an anticipatory effect of Spanish on English productions which results in convergence in adult-directed but not infant-directed speech, though it is possible that both registers will ultimately result in a null effect.

This discrepancy between the two speech registers leads to an interesting result that must be reconciled, as it implies that such anticipation may not be purely automatic. Here I present several possibilities to explain the difference between adult-directed and infant-directed speech.

Hypothesis I: active attempt to provide more extreme cues to language identity in infant-directed speech

One hypothesis is that this anticipatory activation at the phonological level is actually under some level of speaker control rather than being a purely automatic process. Under this hypothesis, when speaking to an infant who is still learning both languages, including phoneme representations and language categories, speakers may attempt to highlight a separation between the two languages. That is, they may reliably produce English words in an English-like manner and Spanish words in a Spanish-like manner, regardless of the position of those words relative to a code-switch. Whereas in adult-directed speech speakers appear to increase activation of a language before completing a lexical switch, such ramping up of activation may not happen in infant-directed speech, in order to produce clearer and more language-canonical tokens in each language. This hypothesis would predict that in infant-directed speech, utterances immediately preceding a language switch would be faster and more fluent and utterances immediately following a language switch would be slower and more disfluent, as speakers need time to manage relative activation of the two languages. It may also predict that code-switches would occur more often inter-sententially, or with a longer pause between languages, in infant-directed speech than in adult-directed speech.

Hypothesis I is in line with other findings in infant-directed speech that suggest some level of hyper-articulation or an increased rate of canonical forms when speaking to infants as compared to speaking to adults. That is, this hypothesis suggests that parents attempt to produce speech that would be easier for their child to perceive, understand, and learn from, even if doing so is slightly more difficult for parents or requires additional effort. However, as discussed in the introduction to this chapter, such patterns are not reliably found across parents, language communities, or acoustic features. Instead, some researchers predict that any observed patterns which may appear to be due to this increased contrastive-ness are in fact unintended by-products of pitch changes in infant-directed speech, or the content of speech to infants. Hypothesis II is based on this proposal.

Hypothesis II: Switching for a different purpose in IDS and ADS

An alternative hypothesis is that code-switching occurs for different pragmatic purposes, or within different syntactic structures, in the two speech registers, and that these content or structural differences may have unintended impacts on fine-grained details of speech production. Several previous studies which investigated the detailed content of code-switching in speech to infants and toddlers (Bail et al., 2015; Kremin et al., 2022) have addressed the relative rates of inter- and intra-sentential code-switching in infant-directed speech, as well as apparent reasons for code-switching. Both found that inter-sentential code-switching was more common than intra-sentential code-switching, but did not compare infant-directed and adult-directed speech. However, they hypothesized that this difference between the two types of code-switches could be driven by children's ease of processing intra-sentential code-switches and a desire to maximize comprehension (at a lexical/semantic level, not necessarily phonological). Further, Kremin et al (2022) suggested that code-switching may serve to teach new vocabulary words and to improve understanding, either by presenting a given word in the language the child knows that word in, or by providing information in both languages to teach translation equivalents.

Code-switching in adult-directed speech, on the other hand, is unlikely to serve these particular purposes, as the listener is a fluent bilingual rather than a language learner. Instead, speakers may be focused on presenting the information in the language that they themselves can most easily access, which could lead to using code-switching as a recovery device for the speaker rather than the listener. They may code-switch in order to emphasize information or signal emotion (Sánchez & Pérez-García, 2020) (though code-switching when emotional may also occur in child-directed speech, i.e. Williams et al., 2020). Perhaps an adult listener's fluency affords the speaker with more freedom to switch fluently, which results in greater integration of activation between the two languages at the phonological level. Further, if intra-sentential code-switching is more common in adult-directed speech than infant-directed speech, then the observed difference between the two speech registers may reflect a process in which phonetic convergence applies within sentences to a greater degree than it does across sentences. Perhaps no evidence of anticipatory convergence was found in the infant-directed speech in both the three-second and five-second time windows because a phrase boundary occurred within those time windows more often in infant-directed speech, and such phrase boundaries inhibited the convergence. A future analysis which characterizes each language switch both in terms of its structural properties (intra-sentential vs. inter-sentential; alternation or single-item insertion) and its apparent purpose (providing translation; consistently naming an item in one language rather than using lexical items from both languages; emotion, etc.) will be necessary to assess whether differences in code-switching structure or purpose are present between the two speech registers, and whether such differences could have led to the observed acoustic outcomes.

Global vs. local effects; language mode vs. code-switching

Despite these discussed local effects of code-switching, particularly in anticipation of a code-switch, the present data show no evidence of more global effects of code-switching. That is, results do not suggest that English VOTs are shorter in code-switched utterances compared to unilingual utterances, or in code-switched stories compared to English-only stories. In fact, even extending the local range of analysis from 3 seconds to 5 seconds somewhat decreased the strength of the local effect of code-switching. This pattern differs somewhat from previously-established results. For example, Fricke et al., (2016) did find an effect of utterance type on the VOT of English voiceless stops, though the effect size was very small (38.7 vs 40.7 ms). Their analysis included 498 tokens from code-switched utterances and 882 from unilingual utterances, whereas the present analysis included 355 and 310, respectively. The null effect observed here could therefore be due to the current underpowered nature of the study. However, they do not have a comparison to our most global analysis (analysis 2), as the corpus was entirely conversations which contain code-switching; for that, we must turn to paradigms involving cued language switching. Such paradigms present a slightly different pattern of results than naturalistic code-switching. While some suggest that the production of code-switched items differs from non-switched items (Goldrick et al., 2014;

Olson 2013), others find an effect of being in a language-mixing context but not of switch trials (i.e. Gavino & Goldrick, 2023). It is unclear, then, whether this observation reflects a true null effect, which would suggest that at least in this task, language mode does not have a progressively gradient effect on the production of English VOTs, or if we simply lacked the power to detect an effect in the present analysis.

Pitch and pitch variability

Analysis 1 was designed to assess the extent to which participants spoke to their infants with a higher and more variable pitch, as is commonly found in infant-directed speech. I did find robust evidence of this effect, thus establishing that mothers (and one grandmother) produced at least a subset of the classically-observed characteristics of infant-directed speech. There is also some indication that English infant-directed speech was overall higher in the code-switching stories than the English-only stories, while English adult-directed speech was overall higher in the English-only stories than the code-switching stories. This interaction is interesting to consider, particularly in light of the methodological choices in segmenting utterances and measuring pitch. As previously discussed, some single words were counted as a unique utterance, while in other cases an utterance was comprised of a long sentence. Single-word utterances were more common in the code-switching stories under this analysis method, as they occurred whenever a single-word insertion was produced (i.e. If somebody said “the gato is black”, “gato” would be counted as a separate utterance. This operationalization of an ‘utterance’ is only true for the analysis of pitch, and not for Analyses 2-4 that focus on VOT). Therefore, if pitch was raised for such single-word insertions, or on code-switched words more generally, we might expect to observe a higher overall pitch in the code-switched stories than in the unilingual stories.

Previous studies support this interpretation: Olson (2012; 2016) found that single-word insertions were produced with greater pitch movement and duration than target words in unilingual sentences. Likewise, English single-word insertions in an otherwise French context were produced with a higher pitch than English unilingual tokens (Muldner et al., 2019). This effect may be stronger in a primarily monolingual language mode than in a more balanced language mode (Olson, 2016), perhaps because a code-switch is less predictable and more effortful for both the speaker and the listener when switching is infrequent. This lack of predictability could lead to hyper-articulation (Bell et al., 2009; Lindblom et al., 1990). However, the authors of those studies used only carefully-constructed read passages and did not analyze the pitch of code-switching in which the speaker controlled their own word choice and sentence structure. Therefore, they primarily addressed issues related to a speaker's articulatory needs, rather than potential communicative or pragmatic aspects of pitch that include listener design. Higher pitch could signal this code-switch and serve to make it more noticeable or distinct from the previous input, or simply as an additional method of attracting attention. Such motivations may be particularly important when talking to infants, leading to the tentative effect observed here in that F0 was greater in code-switched stories than unilingual stories in infant-directed speech only.

However, the confound between annotating single-word utterances and story type in the present analysis could have led to a spurious effect. Stress is associated with pitch, such that stressed syllables tend to be produced with a higher pitch than unstressed syllables (i.e. Fuchs, 2016; Laver, 1994; Plag et al., 2011 for overview). As single-word insertions are more likely to be content words, which often have primary and potentially secondary stress, than function words which are not necessarily stressed, code-switching may result in a higher average F0 of each utterance not because of the code-switching *per se*, but because some utterances (especially in code-switched stories) did not contain unstressed syllables that could bring down the average F0. This possibility could be addressed more directly by measuring the average F0 of every word in both the English-only and the code-switching stories, rather than over utterances of variable length. We could then assess whether words have higher pitch in the code-switching than in the English-only stories, and whether this effect differs between speech registers. It would also be important to focus on the pitch of words within the code-switching stories, using a similar global to local analysis procedure as in the analyses of VOT presented here. If speakers use pitch to highlight code-switching, we would expect that individual words which occur close to the point of a code-switch would differ in pitch from those that occur in unilingual utterances, particularly in infant-directed speech.

Effects of speech rate and word length on VOT

There was a robust effect of word length on VOT, such that words with a longer duration were produced with longer voice onset times. However, after controlling for speech rate by dividing the VOT by the length of the vowel or approximant that followed it, words with a longer total duration were produced with shorter VOTs (relative to the following vowel). The relationship between VOT, speech rate, and word duration is complex (i.e. Boucher, 2002; Kessinger & Blumstein, 1998; Miller et al., 2009), as exemplified by this finding.

To better understand the relationship in this dataset, additional linear mixed effects models were fit which included all 1941 English tokens (across all story types and registers) and only included fixed effects of word length and vowel length, as well as random intercepts of participant and phoneme. Word length, vowel length, and VOT were expressed in terms of milliseconds and all variables (including consonant-vowel ratio) were log-transformed. Using VOT as the dependent variable, there were significant positive effects of both vowel length and word length such that when one variable was held constant, increasing the length of the other was associated with longer VOTs. Further, a model which contained word length and vowel length was a significantly better fit than a model that contained only word length (such as the models used in Analyses 2-4).

Using consonant-vowel ratio as the dependent variable, when only word length was included in the model as a predictor, there was a significant negative effect of word length such that longer words were produced with shorter consonant-vowel ratios. This pattern matches that which was seen in analyses 2-4. However, when vowel length was added to the model there was a significant negative effect of vowel length and a significant *positive* effect of word length. That is, when vowel length was held constant, longer words were produced with *longer* consonant-vowel ratios. However, tokens with longer vowels were produced with shorter consonant-vowel ratios. This model including both word length and vowel length was a significantly better fit that the model only including word length, which may suggest that the models in Analyses 2-4 should have included both word and vowel length as predictors.

To ensure that the exclusion of vowel length in the models did not impact the observed pattern of results, several models in Analysis 4 were re-fit including both word length and vowel length (adult-directed anticipatory, adult-directed perseveratory, infant-directed anticipatory, infant-directed perseveratory). The overall pattern of results was unchanged: there was a significant positive effect of distance for the adult-directed anticipatory subset of data, such that tokens produced farther from a code-switch had both longer VOTs and greater consonant-vowel ratios, and no effect of distance on VOT or consonant-vowel ratio for the other three subsets of data. Furthermore, there was only a significant effect of word length on consonant-vowel ratio for one of the four subsets (infant-directed anticipatory), and that effect was positive, such that longer words were associated with longer consonant-vowel ratios when controlling for vowel length.

A more detailed investigation of the relationship between VOT, vowel length, consonant-vowel ratio, and word length is outside the scope of this dissertation; however, this supplementary investigation suggests that the crucial observed results in analysis 4 were not influenced by a spurious effect or an unintentional oversight in included predictor variables.

One additional consideration regarding vowel length is the inclusion of tokens which were followed by a liquid such as /l/ or /ɹ/. For these tokens, the analysis of consonant-vowel ratio presented in Analyses 2-4 used the length of that liquid as the following "vowel", rather than the vowel following that liquid consonant. For example, for the word "please" (/pliz/), the length of /l/ was used instead of the length of /i/. Follow-up analyses using consonant-vowel ratio as the dependent variable and both vowel length and word length as predictors (along with distance to the code-switch) but replacing the liquid with the true following vowel (/i/ in the example above) were conducted for the four sub-comparisons in analysis 4 (adult-directed anticipatory, adult-directed perseverative, infant-directed anticipatory, infant-directed perseverative). There were 59 such tokens across these four sub-comparisons. In the adult-directed anticipatory case, 9 of the 91 tokens included a liquid immediately following the target phoneme. Replacing the length of those liquids with the length of the following vowel resulted in a marginally significant positive effect of distance to the code-switch ($t(75.83) = 1.919$, $\beta = 0.118$, $p = 0.059$) such that English VOTs decreased as the distance to the code-switch decreased, but this effect did not quite reach significance. While this supplemental analysis suggests that the tokens including a liquid immediately following the target phoneme may have had a small influence on the data, it is likely that with a larger and more highly-powered sample, any variability driven by the acoustics of such tokens would be minimized. In the remaining sub-comparisons there was no significant effect of distance to the code-switch on consonant-vowel ratio, which remained true after replacing the length of the liquid with the length of the true following vowel.

Figure 35: Production study - Analysis 4: Consonant-vowel ratio in code-switched utterances preceding Spanish, liquids replaced with following consonant

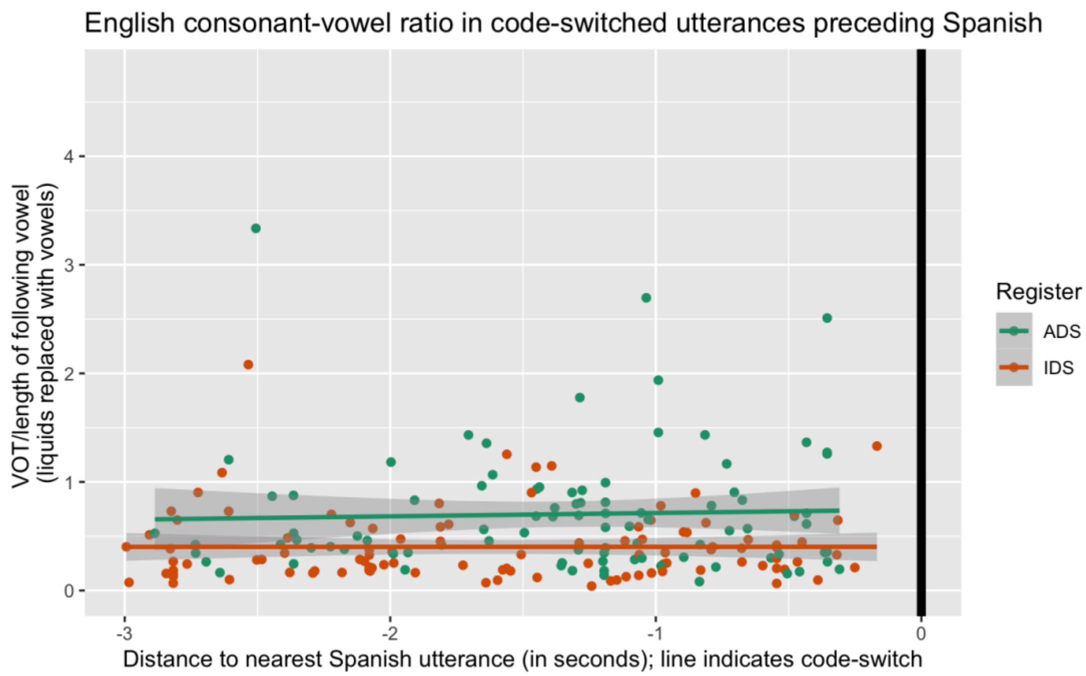
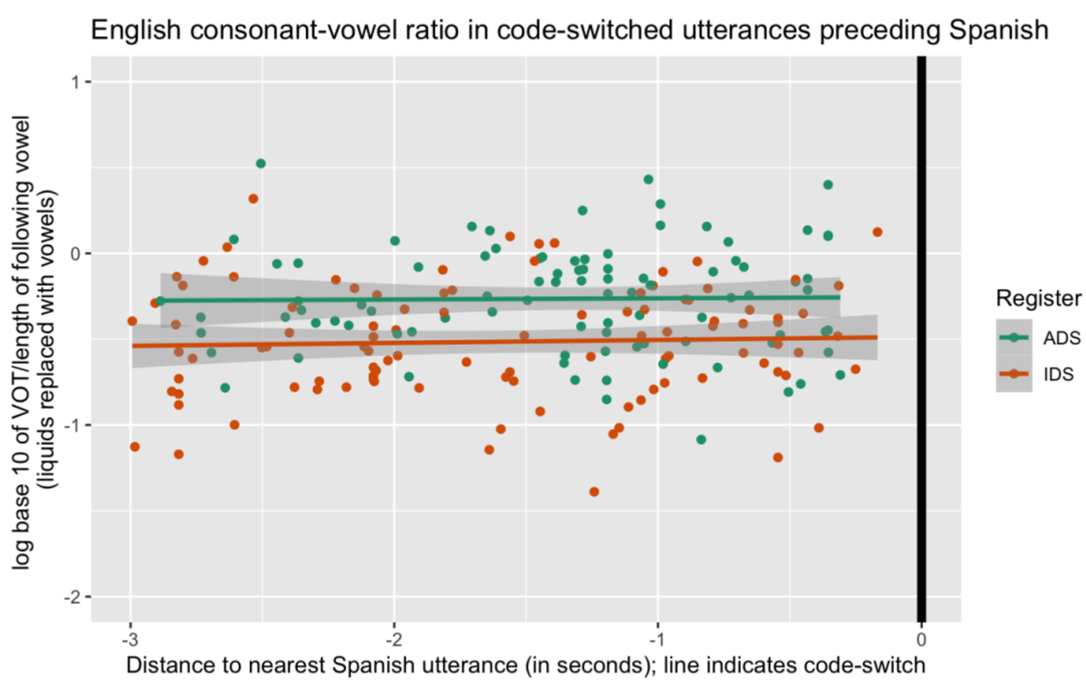


Figure 36: Production study - Analysis 4: log10-transformed consonant-vowel ratio in code-switched utterances preceding Spanish, liquids replaced with following consonant



Additional limitations

As discussed, the current results are based on data from 10 participants. A pilot analysis established that 20 participants would be necessary to obtain 80% power to detect a result in the analysis of code-switching utterances (analysis 4), and recruitment is ongoing in order to reach this target number of participants. Recruiting for this study was unexpectedly challenging. The initial recruitment plan was to target local preschools, churches, and community groups or programs which serve Spanish-English bilingual families with young children. Most organizations which fit these criteria, both in the DC Metro area and in other parts of the United States, either did not respond to our requests or said that they were uninterested. None consented for us to visit the organization in person, regardless of whether we proposed to share information about the study or to simply talk to staff members or families about bilingual language development. A few organizations (< 5) did agree to share our flyer with families or members, but we do not know if this happened. As far as we know no families completed the interest form after receiving a copy of the flyer from their child's preschool or a community group to which they belonged. Hanging physical flyers in neighborhoods in Washington DC which have a high Spanish-speaking population was also entirely unsuccessful.

More successful methods of recruitment involved using Prolific and Lookit/Children Helping Science to identify potential participants who met basic demographic criteria, paid ads on Meta platforms (Facebook & Instagram) targeting women who spoke Spanish and had a child under 2, as well as word of mouth. Although these methods, particularly paid ads, allowed us to reach large numbers of potential participants, attrition still caused difficulty. Further, both Lookit and Meta ads resulted in the interest survey reaching individuals who appeared to meet the demographic criteria for the study based on their answers to the screening questions, but later behaved strangely and were suspected to be would-be scammers. This problem with suspected scammers is a known issue for the Lookit platform; a group of individuals actively try to participate in studies for which they are not eligible with the goal of receiving a gift card, and sometimes appear to engage in this activity simply to waste experimenters' time even in the absence of a monetary incentive. The paid ads on Meta platforms also opened the survey up to a larger group of people and increased the risk of people attempting to participate in bad faith or with minimal attention. Many responses on the screening survey from this group also included people who did not meet the basic criteria for participation which were clearly stated in the ad, such as being able to speak both English and Spanish, being at least 18 years of age, being female, and having a baby.

The attrition problem was perhaps a greater barrier to completion, as participants who did meet all criteria for participating in the study, appeared to be genuine, and claimed to be interested would then fail to complete the consent form and eventually stop responding to the experimenter. Others did complete the consent form and received all materials for the study, but then never completed it despite repeated follow-up emails or text messages. Some individuals repeatedly claimed that they were going to participate soon, but then never did. While many who participated did so within a week of expressing interest and receiving the materials, others did so only with regular reminders. No participant asked to meet with the experimenter to talk more about the study or ask questions about the materials or procedure, despite all being offered that option. Only one potential participant asked for the experimenter to print copies of all the books and mail them to her house; she has yet to participate. It is unclear why so many participants failed to completed the study and stopped responding to follow-up messages, though the independent effort required to fulfill all components of the study likely played a role.

This difficulty in recruitment led to an expansion of participation criteria from the original plan. Namely, one participant over the age of 50 (69; a grandmother rather than a mother) was included, as were 3 with children over the age of 14 months (15, 18, 20 months). In planned variability in participant demographics, individuals who speak a range of varieties of Spanish, and whose relative dominance in English and Spanish varies widely, were included. Of the 10 participants discussed here, 4 are English-dominant, 3 are Spanish-dominant, and 3 have relatively balanced dominance in the two languages. This dominance categorization was based on a combination of their proficiency, age of acquisition, presence of a non-native accent, and proportion of time spent using each language, such that two participants considered in the same dominance category could actually exhibit very different language profiles. The degree to which participants reported engaging in code-switching, both with their infant and with other adults in their language community, did not contribute to this dominance categorization but varies across individuals and language dominance categories.

With only 10 participants, it is difficult to assess how these individual differences in demographic characteristics may have influenced the primary effects of interest considered here. Individual differences were thus not considered in the analyses presented above. However, it is possible, if not likely, that individuals from different language communities or who use their languages in different ways may vary in the detail of how they produce code-switching. Properties of infant-directed speech may also differ based on the age of the child being spoken to (i.e. Cox et al., 2023). Once the full sample has been collected, the child's age, the participant's proficiency in each language or overall language dominance, and the proportion of time the participant reports using code-switching will all be included as potential covariates in all analyses.

An additional limitation of this work is the confound between story identity and condition. That is, all participants were told to tell the "Calli makes a pizza" story to their baby using just English, to tell the "Carmen's Dream" story to the adult while switching between English and Spanish, and so on for all six stories. While some objects are repeated between multiple stories (i.e. cat/kitty) and many actions were described using the same word in multiple stories (i.e. take, tell, carry, call, climb), others only occurred in one particular story (i.e. clown, kangaroo, popcorn, pizza, pepper, taxi). In most cases, we are therefore unable to compare the production of an identical word in multiple registers or language contexts.

Perhaps more crucially, this design choice precludes us from definitively analyzing whether participants structure code-switching differently in adult-directed and infant-directed speech or do so at different rates. No formal analysis of the syntactic structure of code-switched utterances or the apparent purpose of code-switches has been conducted using this data, and impressionistic data based on listening to each story several times does not suggest consistent differences in the rate or type of code-switching between infant-directed and adult-directed speech. However, if participants were to consistently code-switch more frequently in adult-directed than infant-directed speech, for example, or to produce more intra-sentential code-switches in adult-directed speech and more inter-sentential code-switches in infant-directed speech, we would be unable to conclude that this effect is driven solely by a degree of audience design. It could be that the structure of the Carmen's Dream (adult-directed) story would be better suited to different types of code-switching than the Peter and Thomas (infant-directed) story. While it is unclear how the structure of one particular wordless picture book could itself lead to code-switching to a greater degree than the structure of a different wordless picture book, this possibility is still important to consider.

An additional limitation of the current work is that by having participants tell stories to fluent Spanish-English bilingual adults and to infants, effects of speaking to infants and of speaking to listeners who are not fluent in both languages are conflated. That is, some of the observed effects in infant-directed speech are likely specifically due to the age of the participant's interlocutor and a goal to maintain infants' attention or bond with them. However, other effects may also apply to adult listeners who do not know both of the languages in question, and are thus language learners despite their age. Indeed, foreigner-directed speech and infant-directed speech overlap in several acoustic characteristics including vowel hyper-articulation and speech rate (Piazza et al., 2022; Uther et al., 2007). Without a third register condition, conclusions regarding the exact function of some of the acoustic differences between adult-directed and infant-directed speech discussed here remain unclear.

The design of this study was naturalistic, in that parents were not given specific prompts to repeat and were participating in a comfortable environment without an experimenter present. However, the recordings are not fully natural and may not reflect the range of reasons that speakers code-switch. Further, the current analysis does not account for the structural position or pragmatic purpose of a given code-switch nor the distribution of those structural positions in infant-directed and adult-directed speech. This concern is particularly important in the infant-directed speech conditions, as previous research suggests that parents might code-switch for pedagogical reasons or to enhance children's attention or understanding (Kremin et al., 2022). Future research may be able to address this concern by annotating the broader context of each code-switch including the apparent purpose of the code-switch and whether it provided a translation equivalent at the word- or phrase-level. As discussed in the introduction to this chapter, children may acquire language by tracking information across several linguistic and extralinguistic dimensions simultaneously (i.e. Hitczenko & Feldman, 2022; Bernstein Ratner, 1984), which could include the language (or languages) used as well as the acoustic detail of speech. Therefore, considering the local context of each instance of a code-switch in infant-directed speech will improve our understanding of the multi-dimensional and -contextual nature of language development.

Further, while some words were unavoidable for each story (i.e. turtle/tortuga, panda/panda, tree/arból, and ball/pelota in the infant-directed code-switching story), participants otherwise could have chosen to use words based on whether they thought their child would understand. Infants ranged in age from 6 to 20 months such that there was likely a large range in their receptive vocabularies. However, their vocabulary knowledge in English and Spanish was not assessed, so we are unable to determine whether parents chose which words to use (or which language to use for a particular character or object) based on children's expected understanding using the current data. By including this information, future research could also assess whether code-switching occurring proximally to known words differs in its frequency, structural form, or acoustic properties from code-switching occurring proximally to unknown words.

Future directions

As discussed, the primary goal for future analysis is to complete the sample and include data from a full set of 20 participants. Doing so will increase the statistical power and the ability to detect a significant effect if one exists. It may also introduce additional variability and make it possible to evaluate potential individual differences. The other immediate goal for this study is to measure the VOTs of all voiceless stops in Spanish and to complete the same analyses discussed above using Spanish productions. Although it is somewhat unlikely, based on previous studies, that we will see an effect of English on the VOT of Spanish voiceless stops (i.e. Bullock et al., 2006), this analysis is still important and will broaden our understanding of the interaction between Spanish and English phonological representations. For example, if there is an anticipatory effect of English on Spanish productions (at least in adult-directed speech), and the current lack of a perseverative effect of Spanish on English productions holds, we could conclude that this lack of Spanish perseveration following a code-switch into English is *caused by* the anticipation of English. Preparing for this switch into English would diminish the necessity for continued activation of Spanish phonology to a degree that influences English productions. Measuring the VOTs in Spanish will also allow for an evaluation of the extent to which parents differentiate the two languages in terms of the distribution of voiceless stop productions.

Though the analyses presented in this study focus on the voice onset times of voiceless stops, I do not believe that this is the only acoustic property that may be influenced by code-switching, nor that it is the only cue listeners might use to anticipate or detect a language switch. Rather, as established in Chapter 2, this initial work is focused on voice onset times of voiceless stops because they have been the primary cue studied in previous work, allowing for comparison between infant-directed speech and the established literature on adult-directed speech, and because VOT is relatively straightforward to measure. Future analyses will assess other potential cues whose relationship to code-switching is less established. The main properties of interest in this future work will be speech rate and disfluencies (Fricke et al., 2016), rhythmic and durational properties, vowel quality, and prosodic contours (Piccinini & Garellek, 2014; Shen et al., 2020). Fricke et al. (2016) found that, in addition to a reduction in VOT in anticipation of a code-switch, speech preceding a code-switch was also slightly slower and had greater rates of disfluency compared to speech in unilingual utterances. We will test whether similar effects also occur in infant-directed speech, which tends to be slower than adult-directed speech even in its unilingual form. Further, Piccinini & Garellek (2014) suggested that Spanish-English bilinguals may produce code-switched sentences with a prosodic contour that is ambiguous between standard contours in English and Spanish. However, that analysis is based on one speaker and may not be representative of broader patterns. The corpus collected through this work will allow us to test for similar prosodic contours in a larger sample of speakers, and to compare such contours in adult-directed and infant-directed speech.

While rhythmic and durational properties of code-switching have not been addressed previously, at least to the author’s knowledge, it is likely that such cues could be particularly meaningful when considering the question of language discrimination and learning. That is, as listeners are sensitive to these cues and able to use them to perceptually discriminate passages of different languages, the existence of a switch in rhythmic or durational patterns at the point of a code-switch may be a uniquely salient cue to that code-switch. However, if as seen in the (adult-directed) English voiceless stops, an upcoming language may be activated at the phonological level before a lexical switch occurs, such rhythmic and durational properties may be ambiguous before the point of a code-switch. In an English-to-Spanish code-switch, though, it may be that English speech has a lower normalized vocalic pairwise variability index than it does in a unilingual utterance, as Spanish tends to have a lower nPVI-V score than English does (Grabe & Low, 2002). While some aspects of these cross-linguistic differences in rhythm are driven by the syllable inventories of different languages, with English using more consonant clusters than Spanish, there exists individual variability which is both idiosyncratic and affected by a speaker's native language. Namely, L1 Spanish – L2 English speakers have a lower nPVI-V score in English than L1 English speakers do, and L1 English – L2 Spanish have a higher nPVI-V score in Spanish than L1 Spanish speakers do (White & Mattys, 2007). Early bilingual adults do, though, show evidence of language-specific rhythm production (Aldrich, 2020). Therefore, we may expect that this global effect of language knowledge and dominance would also lead to a local effect of language activation on rhythmic characteristics.

Vowel characteristics likewise have not been studied in naturalistic code-switching. However, Gavino and Goldrick (2023) found that in a cued-switching paradigm, being in a mixing context (i.e. a picture naming session in which participants were occasionally cued to switch languages) led to a decreased contrast in the heights of Spanish /i/ and English /i/ and Spanish /e/ and English /ɛ/, as well as the backing of Spanish /e/ and English /e/, relative to being in a single-language context. Further, Muldner et al. (2019) found tentative evidence that French vowel quality when reading otherwise-English sentences differed from French vowel quality in entirely-French passages. While cued-switching and sentence-reading paradigms are not directly comparable to naturalistic switching as they remove speakers' agency and limit the potential for speech planning, these patterns does provide some evidence that vowel quality may be impacted by the relative activation of a bilingual's two languages. Future research will investigate similar effects of switching on vowel quality, but in this natural speech context.

It could also be that individual words, or words that have particular properties, have a greater potential for cross-language phonetic transfer than other words. Under a lexical triggering account of code-switching, cognates may facilitate a code-switch because cognates result in high degrees of cross-language activation (i.e. Broersma et al., 2020; Broersma & De Bot, 2006; Clyne, 1980). Using a cognate is correlated with code-switching in the same or a shortly-upcoming clause. It is possible that this facilitation of code-switching from cognates could also extend to effects of code-switch anticipation, such that VOTs would converge more with the other language when they occur in a word that has more phonological similarity with the other language. That is, in an English-Spanish code-switched sentence that has either the word "peanut" or "pitcher" immediately preceding the code-switch, perhaps the /p/ in "peanut" would be produced with a shorter VOT than the /p/ in "pitcher" because the tense /i/ in "peanut" does exist in the Spanish inventory but the lax /ɪ/ in "pitcher" does not. Crucially, a finding in support of this prediction would not show that phonological similarity between the vowels was a trigger for the code-switch occurring after the target word, because the change in VOT would occur even *before* the vowel was produced. However, it would suggest that the cross-language phonetic transfer discussed here is driven by activation of the two target languages at the phonological level, such that transfer occurs to a greater extent when there is more cross-language activation.

Chapter 6: Overall Discussion and Conclusion

Summary of results

Perception approach: adults

In Experiment 1, monolingual and bilingual adults listened to a French-Mandarin bilingual tell stories while producing inter-sentential code-switches between the two languages. They continually identified which language he was speaking. I found that bilingual adults (more specifically those who hear code-switching more frequently) were faster to identify a code-switch when the story was told with naturalistic acoustic cues to code-switching than when such natural cues were absent. Monolingual adults, and bilingual adults with minimal experience listening to code-switching, showed no such effect of condition on their accuracy to identify a code-switch, though all groups had the same overall reaction time across conditions. Experiment 2 expanded this finding by showing that anticipatory acoustic cues to a code-switch, but not perseverative cues, were particularly helpful in allowing bilingual adults to detect it. Overall, as discussed extensively in chapter 3, these experiments suggest that bilingual adults who have experience with code-switching and some pre-existing representation of both French and Mandarin Chinese attend to sub-phonemic detail of the two languages. Acoustic properties, particularly of Mandarin, may change in a consistent and perceptible manner preceding a Mandarin-to-French code-switch such that listeners can better predict and respond to such a switch.

Perception approach: children and infants

In Experiment 3, monolingual children completed the same experiment as the adults in Experiment 1. However, unlike the adults, children did not show an overall ability to identify French and Mandarin Chinese while listening to a single bilingual speaker tell stories that contained code-switching between the two languages. Their performance was at chance both when hearing naturalistic acoustic cues and unnatural acoustic cues to code-switching. This lack of an overall ability could be driven by several sources including attention, knowledge of each language, understanding of the task, and meta-linguistic ability. However, this experiment did not show any evidence that the presence of natural acoustic cues to code-switching influenced children's performance. This lack of an effect can only be tentatively interpreted given the current results, but if true, suggests that children either cannot perceive differences in the acoustic properties of French and Mandarin that do or do not signal a code-switch, or that they can perceive such cues but that the information does not affect their higher-level processing and prediction. In Experiment 4, monolingual infants also continually categorized French and Mandarin Chinese. Yield rates were low, impeding the ability to meaningfully interpret the results. However, a small number of infants could discriminate French and Mandarin Chinese, including identifying a code-switch between the two languages. Due to sample sizes we cannot conclude whether this ability is influenced by the presence of naturalistic acoustic cues to code-switching; only children who heard unnatural acoustic cues provided enough data that their responses can truly be interpreted.

Production approach

In the analysis of production data, in which 10 Spanish-English bilingual women narrated wordless picture books in infant-directed and adult-directed speech, there was local effect of code-switching on English voice onset times in adult-directed speech only. Namely, mothers produced English VOTs which were shorter (more Spanish-like, but still within English norms) immediately preceding a code-switch into Spanish than when English voiceless stop tokens preceded a code-switch by a longer latency. I therefore replicated a previous finding regarding anticipatory effects of English-Spanish code-switching in adult-directed speech. This pattern was not seen in infant-directed speech, or following a Spanish-to-English code-switch in either speech register. This finding suggests that parents may be producing speech which is fundamentally different, and perhaps makes more of a distinction between the two languages, when speaking to their infants as compared to speaking to fluent adults.

Consequences for bilingual language development

One of the key concerns which motivated this work was a question of how exposure to code-switching might influence bilingual language development. Specifically, I was interested in how acoustic properties of code-switching directed to children might make it easier (or harder) for them to distinguish between two languages while listening such that they could separately track distributional information about each linguistic code. I predicted that natural acoustic cues to code-switching might make it easier to detect a code-switch for listeners who have some existing representation of each language or expectation about the properties of code-switching; however, these same cues might hurt identification of a code-switch for listeners with weaker representations who would rely more on hearing a clear difference between the languages. The first part of this hypothesis was upheld; Spanish-English bilingual adults who hear code-switching frequently were faster to detect a code-switch when they heard natural acoustic cues to code-switching, especially when those cues occurred before the code-switch. The second part, however, was not: there is no evidence that these same natural cues to code-switching *hurt* detection of a language switch in English monolingual adults, in bilingual adults with minimal code-switching experience, or in monolingual children. There is not enough data to address this question in English monolingual infants.

In fact, the production study actually suggests that infants growing up in bilingual homes may not hear these natural acoustic cues to code-switching directed to them, even if their mother produces anticipatory effects of Spanish on English when speaking to adults. That is, even if it is true that natural acoustic cues to code-switching (which could effectively blur the difference between two languages) make it more difficult for infants or young children to discriminate the languages, such an effect may be irrelevant in terms of children's true linguistic experience and therefore not have an opportunity to impact language development and learning. If instead, as the production study suggests, the acoustic detail of English produced temporally close to Spanish does not differ from English produced more temporally distant from Spanish in IDS, including in an English-only interaction, then perhaps the "unnatural" condition in the perception studies is a more apt comparison to the speech directed to bilingual infants than the "naturalistic" condition is. In Experiment 4, several infants in this unnatural condition did show an ability to discriminate between French and Mandarin and to associate each language with a physical direction; they looked in the correct direction following a code-switch. It is possible that this discrimination ability would be even stronger if bilingual infants participated in a similar task but listened to the two languages to which they are exposed, as they already have existing representations of each language to draw on. However, the recent study by Birulés et al. (2024), which found that monolingual infants exposed to Spanish or Catalan, but not bilingual infants exposed to both languages, noticed when an audiovisual speech sample switched between these two languages conflicts with such a prediction.

Though the experiments in this dissertation were not designed to specifically test whether code-switching impacts children's language development, this combination of results does have several implications for bilingual language development. First, the finding from Experiment 4 that several infants could identify French-Mandarin code-switches at above-chance levels suggests that even children with no prior exposure to code-switching can find patterns in speech and produce differential behaviors in response to those patterns. This skill would presumably be necessary to learn two languages in a bilingual environment and keep the input data for them distinct. However, extending those patterns to conceptual categories of a "language" and enacting meta-linguistic judgments on such language categories, as was required in the task presented to the children and adults, is a more abstract behavior that may require more cognitive development or linguistic awareness than is available to 5-8 year-old monolingual children. These results correlate with previous findings of language differentiation in infants and young children reviewed in Chapter 1. The phonological characteristics of bilingual infants' babbling behavior differs based on the language spoken by their interlocutor (Sundara et al., 2020), and older bilingual children show some ability to change which language(s) they produce based on the language knowledge of their interlocutor (Genesee et al., 1995; 1996; Comeau et al., 2003; 2007). However, whether these changes are conscious or a more low-level form of pattern matching is unclear. Patterns of phonological development in bilinguals suggest that languages become more distinct as children develop, which could reflect increasing conscious awareness of the languages as categories themselves, along with improved motor control of the articulators.

Future research that connects a parent's speech patterns with their child's language development will be needed to assess the relationship between *hearing* acoustic cues to code-switching and *producing* indicators of language differentiation in early childhood.

Considerations of phonetic detail of the language pairs

In the perception approach (Experiments 1-4), I used French and Mandarin as example languages. As discussed in Chapter 2, this language pair was chosen for a combination of practical and theoretical reasons. However, the questions assessed in this work are not specific to French and Mandarin. Instead, these results are meant to showcase how natural acoustic cues to code-switching may impact identification of a code-switch for listeners with minimal semantic knowledge of the two languages in question. It is likely, though, that language pairs differ regarding the extent to which such natural cues to code-switching are present in various acoustic dimensions. For example, whereas French-Mandarin and Spanish-English are language pairs that do not align in the distribution of voicing of stop consonants, French-Spanish and Mandarin-English *do* align. Therefore, anticipatory convergence in the voicing of stop consonants could not occur prior to a code-switch in the French-Spanish or Mandarin-

English-like language pairs, and would not help listeners to anticipate an upcoming code-switch. It is possible, though, that convergence in other acoustic dimensions such as vowel quality, speech rate, or intonation patterns may occur in language pairs such as French-Spanish or Mandarin-English. Further, while the expectation is that VOT was at least one of the cues that was present in the French-Mandarin code-switching, this is not necessarily true. Acoustic analyses of the French-Mandarin stimuli used in Experiments 1-4 will help to elucidate the exact anticipatory cues that participants may have been attending to in the present experiments and to establish cues that could be assessed in other language pairs in future perceptual studies.

In no way does the work presented here suggest that acoustic cues to code-switching are the only, or even necessarily primary, influence on listeners' ability to identify a code-switch. On the contrary, monolingual and bilingual adults in experiments 1 and 2 succeeded at identifying a switch between Mandarin and French in just over two seconds, whether those cues were present or absent. Hearing (particularly anticipatory) cues to code-switching resulted in a small boost to reaction time for (at least some) bilinguals, but clearly such cues were unnecessary for overall language discrimination. There are many other linguistic and extra-linguistic differences between pairs of languages that could support this discrimination, whether those differences are subject to cross-language phonetic transfer at a code-switch or not.

Rhythm, or durational variability, for example, is commonly discussed in this literature, and may be particularly salient and noticeable given that it applies over phrases or longer sections of speech rather than individual phonemes. The durational properties that are most relevant in discriminating a given language pair may be driven by the detail of each of the languages, and therefore differ between language pairs, such that the more salient or reliable cues differentiating a pair of languages are weighted more strongly (White et al., 2012). Testing a variety of language pairs which range in their relative VOT, vowel quality, speech rate, various durational properties, and intonational similarity may be necessary to assess how listeners weight the importance or usefulness of each dimension, including convergence in that dimension, when identifying a language switch.

Broader conceptualizations of "code switching"; language mode

As established in Chapter 1, this dissertation uses the term "code-switching" to refer to bilingual language switching, in which a bilingual may use lexical items from multiple languages in a single or subsequent utterances. This code-switching is a rule-governed behavior; bilinguals switch languages at points when the syntactic structure of the two languages overlaps (Poplack, 1980). While in some cases, i.e. of switches involving alternation, the grammar of the two languages may also switch, other cases may be thought of more as an insertion of a single lexical item or short phrase from an embedded language into a matrix language (Muysken, 2000; Myers-Scotton, 1997).

However, the term code-switching has in recent years expanded in meaning beyond its use in linguistics to refer more broadly to the ways in which individuals may change their communication patterns or manner of presenting themselves to different audiences or in different contexts. This broader conceptualization of code-switching incorporates aspects of speech register and audience design and is related to Grosjean's discussion of bilingual language modes (i.e. 2001, among other writings). Language mode exists on a continuum and could range from a monolingual mode, in which a bilingual is in a situation that only licits the use of one of their languages and the illicit language is thus relatively inactive, to a bilingual mode, in which both languages are active and could be used. Intermediate positions on this continuum may result from situations in which all individuals in an interaction know both languages but the broader situation (topic, location, etc.) calls for the use of only one of the languages. In this sense, the language mode as determined by interlocutors' language knowledge, physical location, and purpose of the interaction may also constitute somewhat of a “switch.”

Though we did not see effects of language mode in the production study discussed in Chapter 5, all participants were in their own homes (where they report using both languages) and speaking to an individual who knows or is learning both languages. The only factor influencing a monolingual language mode in the four single-language stories was the externally-determined experimental instructions. Perhaps this manipulation was not sufficient to override other more internal indicators of the language mode. While the production study did aim to address the effect of language mode (unilingual vs. code-switching) on voiceless stop production, it did not take a more focused look at how the broader social and linguistic context may influence an individual’s word choice, morphosyntax or even other phonetic aspects of production. However, a socio-phonetic examination of code-switching that addresses how phonetic properties of code-switching vary across social contexts would extend this work in important directions.

More broadly, the pattern whereby a code-switch at the lexical level may not be perfectly aligned with phonetic realizations of two languages requires us to clarify our understanding of when (and where) a code-switch occurs. Indeed, as established in the language mode literature, changes that may occur at physical, cultural, or pragmatic levels have the capacity to impact a listener or speaker’s expectations at the syntactic, lexical, or phonological levels. Even if code-switching does not occur in a given interaction, the increased likelihood of code-switching might be thought of as a “switch” itself. However, the current research does not necessarily imply that acoustic changes that may occur on a different timeline from lexical ones should be considered as a separate or true “switch point”, as such acoustic changes tend to be sub-phonemic and to have a wide range of variability. A given single phone cannot be categorized as having derived from one of two languages with as much success as a given single lexical item.

Potential clinical implications

This dissertation is focused on typically developing children and adults, rather than considering populations with hearing, speech, or language disorders. However, an exploration of the properties of language input and perceptual abilities that support bilingual language development is perhaps more important for such children than it is for those who are typically developing. Bilingual parents hear advice from a variety of sources regarding whether their children should be exposed to multiple languages, including some advice which claims that children with cognitive, language, or hearing disorders should focus on English as the language of instruction (Kremer-Sadlik, 2005; Trelles & Castro, 2019). Parents whose children who are receiving services from an SLP may be particularly likely to follow these single-language recommendations and limit the extent to which their child hears the family language (Blanc et al., 2022). Following this recommendation can prevent the child from participating in family activities and communicating with older family members who do not speak English.

However, there is limited research to support the idea that bilingual language exposure itself impedes children's ability to develop vocabulary and other linguistic knowledge. In fact, bilingual children with autism spectrum disorder score similarly to their monolingual peers, if not better, in measures of cognitive function, language, vocabulary, and adaptive functioning (M. Wang et al., 2018). Bilingual children with hearing loss who received intervention services in both of their languages outperformed their peers receiving English-only services in their English expressive language and auditory comprehension abilities (Bunta et al., 2016). Bilingual children with hearing loss also learn English words equally well regardless of whether those words are taught only in English or in both English and Spanish (McDaniel et al., 2019). Further, bilingual children with Down Syndrome show comparable language abilities to their monolingual peers, including in expressive vocabulary (Kay-Raining Bird et al., 2005), phonological awareness (Ward & Sanoudaki, 2021), and an ability to use syntactic bootstrapping to learn word meanings (Cleave et al., 2014). However, these studies primarily focused on overall language outcomes. They did not characterize the process by which monolingual and bilingual children with hearing or language disorders undergo language acquisition and development, and tended not to consider potential effects of code-switching.

Children with hearing loss

One potential fruitful area of future research will be to assess the phonetic detail of bilingual speech directed to infants and young children who have hearing loss. As discussed in the introduction, the pitch properties of infant-directed speech to children with hearing loss may correspond to their hearing age rather than chronological age (Bergeson et al., 2006; though differences are. not always found as in Lovcevic et al., 2020). Vowel space may not be affected by a child's hearing abilities, though (Wieland et al., 2015). While both infants with both hearing aids and cochlear implants (Robertson et al., 2013; Segal & Kishon-Rabin, 2011; Y. Wang, Bergeson, et al., 2018; Y. Wang et al., 2017) appear to prefer infant-directed speech over adult-directed speech or other types of noise to a similar extent as infants with normal hearing, such patterns only show that infants can detect a difference between the two registers and that IDS may be better suited to grab and maintain their attention. They do not tell us how infants and young children with hearing loss develop language using the information that they hear, let alone how they represent that language. The exact acoustic detail of the speech available to infants, and the nature of the representations learned from that input, must depend on both infants' hearing ability and the properties of their parents' speech.

This future work should include detailed examinations of children's hearing ability, including their baseline audiological abilities (without amplification), the processing algorithms imposed by their amplification system(s), and their audiological and speech discrimination abilities with amplification. This work should then measure the manner and degree to which the child's parents differentiate languages, both during a code-switched utterance and more globally. If this differentiation occurs in an acoustic dimension to which the child has audiological access (i.e. is temporal in nature, is sufficiently loud, involves a frequency difference that the child can hear, etc.), it would suggest that children with hearing loss may be able to make use of such acoustic cues to distinguish their two languages. If, however, no dimension of differentiation occurs to an extent that a child could perceive it, then children must rely on extra-auditory cues to differentiate languages. For example, a child could attend to the contexts in which individual lexical items co-occur, determine the language category to which each lexical item belongs, and use that language tag at the lexical level to differentiate languages. This, admittedly, is a bit of a circular proposition. Further, as discussed in chapter one, the extent to which bilingual adults (with normal hearing) truly need to access a lexical item's language category membership in order to comprehend that word may be minimal.

Note that a finding that parents do not produce differentiation that children with hearing loss can easily hear would not suggest that parents should not expose their child to multiple languages, or even necessarily that the parent should restrict code-switching in the child's presence. It would instead provide an explanation regarding why that particular child may be delayed in learning language-specific patterns, if such a delay exists. It would also suggest that the dimensions along which children's language developmental and speech perceptual systems track distributional information include some which are non-auditory. In terms of intervention, it may suggest that the child would need extra assistance or explicit training learning to attend to those cues which are difficult but possible for them to access. It may also provide additional information for manufacturers of hearing aids or cochlear implants regarding the processing algorithms that should be used. Many of these algorithms are designed based on the acoustic properties of English and other Western languages, and are not necessarily the most useful or preferred for users who speak other languages (see Chasin, 2008; Matsumoto et al., 2018); tuning the algorithms based on the specific languages that a child speaks may improve their outcomes.

Children with language or cognitive disorders

For children with language or cognitive disorders, code-switching or bilingual language input may present an added difficulty not because it is difficult for such children to access and perceive cues which do exist in the speech signal, but because they may struggle to attend to the correct information and appropriately store it in memory. In the broad model of language development assumed here, children must track information from a variety of sources simultaneously. For each piece of information (speech sound token, word, linguistic code, etc.) they also need to track how often it occurs in other, potentially relevant, linguistic or non-linguistic contexts and store such co-occurrence patterns in memory. This distributional information needs to be accessible such that children's "compare and contrast" processing algorithms might propose and adjudicate between several hypotheses regarding more abstract linguistic patterns and structures. Children who have language or cognitive disorders may be unable to attend to the requisite number of features to detect complex patterns, or may not detect such complex patterns as easily as typically-developing children do (see Goffman & Gerken, 2023 for a discussion of a theory of DLD involving deficits in sequential pattern learning). Perhaps for these children, tracking information for multiple languages simultaneously is easier when those languages differ among a greater number of features or when differences among multiple features occur at the same time to limit the number of conditions that must be specified in a given pattern. This proposal is clearly speculative, but could make predictions regarding the learnability of new words in a code-switched context for bilingual children with developmental language disorder: performance may be better when the phonological distinction between the two languages is enhanced and when code-switching occurs in a predictable syntactic structure.

Conclusion

Overall, this dissertation reached its ultimate goals in combining evidence from speech production and speech perception regarding the usefulness of acoustic cues to code-switching in allowing listeners to detect a code-switch. I found that anticipatory phonetic cues to code-switching are present in speech to adults *and* that adults who have some familiarity with code-switching can use those cues to detect a language switch even in the absence of comprehension. However, listeners who are less accustomed to hearing code-switching may not be able to use such cues. In speech to infants, on the other hand, I did not find evidence that anticipatory or perseverative phonetic cues to code-switching are present. Although experiments 3 and 4, which focused on how children and infants respectively perceive naturalistic cues to code-switching in languages they do not know, resulted in somewhat inconclusive results, they do not provide evidence that children would find it more difficult to detect code-switching when hearing speech that does *or* does not include phonetic cues to code-switching.

From a methodological perspective, I believe that the approach set forth in this dissertation, namely one of addressing the speech that infants hear combined with a cross-sectional investigation of speech perception, is one that can be applied to other topics and research questions to provide a more complete picture of language learning and development. Understanding infants' speech perception abilities is only informative if the sounds that they can discriminate or representations they appear to have reflect the reality of their linguistic environments. Likewise, highlighting some linguistic structure in speech to children is only useful if they can both attend to and learn from that acoustic cue. The cross-sectional design of the present approach was faster and potentially less resource-intensive than a longitudinal design with similar goals would have been, while still providing crucial information.

Appendices

Appendix I: Experiment 1 Stimuli

Audio stimuli are available on OSF at: https://osf.io/479ry/

Participants were assigned to one of four lists.

All participants heard all six stories (Chicken, Donkey part 1, Donkey part 2, Khalai, Nozibele part 1, Nozibele part 2) twice (once in the French-Mandarin order and once in the Mandarin-French order).

In lists 1 and 3, the French-Mandarin order of each story was presented in the naturalistic condition and the Mandarin-French order of each story was presented in the unnatural condition.
In lists 2 and 4, the French-Mandarin order of each story was presented in the unnatural condition and the Mandarin-French order of each story was presented in the naturalistic condition.

In lists 1 and 2, Mandarin was on the left and French was on the right.
In lists 3 and 4, French was on the left and Mandarin was on the right.

Appendix II: Experiment 1 Model Specifications

RT full model (language group)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
runid	(Intercept)	0.0663542	0.25759			
	lang	0.0003521	0.01877		-0.7	
story	(Intercept)	0.0100905	0.10045			
	cond	0.0048785	0.06985		-0.12	
	lang	0.0024208	0.0492		0.17	-0.01
Residual		0.1101387	0.33187			

Fixed effects:

	Estimate	Std. Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.998	0.1281	0.746924	1.249076	82.29	7.788	1.80E-11
lang	0.00251	0.01637	-0.0295752	0.0345952	11.24	0.153	0.8809
cond	0.005579	0.02166	-0.0368746	0.0480326	10.5	0.258	0.8017
grp	-0.01492	0.03011	-0.0739356	0.0440956	74.82	-0.495	0.6217
ManCent	-0.1042	0.06844	-0.2383424	0.0299424	75.2	-1.522	0.1321
FrCent	0.06728	0.04915	-0.029054	0.163614	75.97	1.369	0.1751
age	-0.007603	0.003811	-0.0150726	-0.0001334	75.43	-1.995	0.0497
lang:cond	0.002872	0.007857	-0.0125277	0.01827172	1611	0.366	0.7148
lang:grp	0.004903	0.00813	-0.0110318	0.0208378	74.02	0.603	0.5483
cond:grp	-0.01422	0.007883	-0.0296707	0.00123068	1595	-1.804	0.0715
lang:cond:grp	-0.002068	0.007854	-0.0174618	0.01332584	1610	-0.263	0.7923

RT simplified model (language group)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.0667651	0.25839		
	lang	0.0002888	0.01699	-0.72	
story	(Intercept)	0.0100684	0.10034		
	cond	0.0048512	0.06965	-0.11	
	lang	0.0021668	0.04655	0.16	0.01
Residual	0.1100447	0.33173			

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.9898	0.128	0.73892	1.24068	84.28	7.736	2.01E-11
cond	0.00557	0.0216	-0.036766	0.047906	10.57	0.258	0.8015
grp	-0.01265	0.02955	-0.070568	0.045268	76.84	-0.428	0.6699
age	-0.007324	0.003804	-0.0147798	0.00013184	77.32	-1.925	0.0579
cond:grp	-0.01419	0.007879	-0.0296328	0.00125284	1597	-1.801	0.072

RT full model (CS frequency)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.0662148	0.25732		
	lang	0.0003665	0.01914	-0.69	
story	(Intercept)	0.0101969	0.10098		
	cond	0.0048601	0.06971	-0.12	
	lang	0.0024251	0.04925	0.18	0
Residual		0.1099298	0.33156		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.9973	0.128	0.74642	1.24818	82.39	7.789	1.78E-11
lang	0.002545	0.01638	-0.0295598	0.0346498	11.24	0.155	0.8793
cond	0.005387	0.02162	-0.0369882	0.0477622	10.52	0.249	0.808
CSCent	0.01287	0.02039	-0.0270944	0.0528344	76.38	0.631	0.5299
ManCent	-0.103	0.06831	-0.2368876	0.0308876	75.24	-1.508	0.1358
FrCent	0.06504	0.04931	-0.0316076	0.1616876	76.28	1.319	0.1911
age	-0.007585	0.003807	-0.0150467	-0.0001233	75.47	-1.993	0.0499
lang:cond	0.002832	0.007849	-0.012552	0.01821604	1611	0.361	0.7183
lang:CSCent	-0.002557	0.005502	-0.0133409	0.00822692	75.63	-0.465	0.6435
cond:CSCent	0.01365	0.005393	0.00307972	0.02422028	1600	2.531	0.0115
lang:cond:CSCent	0.00036	0.005366	-0.0101574	0.01087736	1614	0.067	0.9465

RT simplified model (CS frequency)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.0665384	0.25795		
	lang	0.0002978	0.01726	-0.72	
story	(Intercept)	0.0101659	0.10083		
	cond	0.0048409	0.06958	-0.11	
	lang	0.002167	0.04655	0.18	0.02
Residual		0.1098279	0.3314		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.9889	0.1277	0.738608	1.239192	84.35	7.741	1.95E-11
cond	0.005343	0.02158	-0.0369538	0.0476398	10.57	0.248	0.8091
CSCent	0.01326	0.01995	-0.025842	0.052362	78.18	0.665	0.5082
age	-0.007298	0.003796	-0.0147382	0.00014216	77.32	-1.922	0.0582
cond:CSCent	0.01364	0.005391	0.00307364	0.02420636	1601	2.531	0.0115

ACC full model (language group)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.109672	0.33117		
	lang	0.007774	0.08817	-0.22	
story	(Intercept)	0.065004	0.25496		
	cond	0.008022	0.08957	-0.51	
	lang	0.003656	0.06046	0.39	-0.77
Residual	0.233004	0.4827			

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.016	0.1787	0.665748	1.366252	83.85	5.688	1.83E-07
lang	-0.02134	0.02309	-0.0665964	0.0239164	15.59	-0.924	0.3695
cond	0.006643	0.02829	-0.0488054	0.0620914	10.64	0.235	0.8188
grp	0.02425	0.03909	-0.0523664	0.1008664	74.53	0.62	0.537
ManCent	0.125	0.08925	-0.04993	0.29993	74.75	1.4	0.1656
FrCent	-0.07882	0.06407	-0.2043972	0.0467572	75.44	-1.23	0.2224
age	0.009365	0.004972	-0.0003801	0.01911012	75.06	1.883	0.0635
lang:cond	-0.01317	0.01144	-0.0355924	0.0092524	1602	-1.152	0.2496
lang:grp	0.006024	0.01511	-0.0235916	0.0356396	70.9	0.399	0.6914
cond:grp	0.01762	0.01146	-0.0048416	0.0400816	1590	1.537	0.1246
lang:cond:grp	-9.648E-05	0.01144	-0.0225189	0.02232592	1601	-0.008	0.9933

ACC simplified model (language group)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.10948	0.33088		
	lang	0.007713	0.08782	-0.24	
story	(Intercept)	0.063783	0.25255		
	cond	0.007276	0.0853	-0.49	
	lang	0.003559	0.05966	0.38	-0.77
Residual		0.233025	0.48273		

Fixed effects:

<table>
<tr><th></th><th>Estimate</th><th>Std.Error</th><th>95%CI-low</th><th>95%CI-high</th><th>df</th><th>t value</th><th>Pr(>|t|)</th></tr>
<tr><td>(Intercept)</td><td>1.032144</td><td>0.174139</td><td>0.69083156</td><td>1.37345644</td><td>88.082673</td><td>5.927</td><td>5.88E-08</td></tr>
<tr><td>age</td><td>0.009085</td><td>0.004927</td><td>-0.0005719</td><td>0.01874192</td><td>77.859473</td><td>1.844</td><td>0.069</td></tr>
</table>

ACC full model (CS frequency)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.109368	0.33071		
	lang	0.007849	0.08859	-0.21	
story	(Intercept)	0.065092	0.25513		
	cond	0.008015	0.08953	-0.51	
	lang	0.003651	0.06042	0.39	-0.77
Residual		0.232814	0.48251		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.016	0.1786	0.665944	1.366056	83.8	5.692	0.00000018
lang	-0.02124	0.0231	-0.066516	0.024036	15.63	-0.919	0.3719
cond	0.006896	0.02828	-0.0485328	0.0623248	10.66	0.244	0.812
CSCent	-0.01991	0.02639	-0.0716344	0.0318144	75.61	-0.754	0.453
ManCent	0.123	0.08911	-0.0516556	0.2976556	74.74	1.38	0.1716
FrCent	-0.07616	0.06425	-0.20209	0.04977	75.6	-1.185	0.2396
age	0.009368	0.004968	-0.0003693	0.01910528	75.04	1.886	0.0632
lang:cond	-0.01317	0.01143	-0.0355728	0.0092328	1602	-1.152	0.2494
lang:CSCent	0.001172	0.01022	-0.0188592	0.0212032	72.21	0.115	0.909
cond:CSCent	-0.01477	0.007852	-0.0301599	0.00061992	1591	-1.881	0.0601
lang:cond:CSCent	0.0002228	0.007801	-0.0150672	0.01551276	1606	0.029	0.9772

RT full model (bilinguals)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0800245	0.28289	
	lang	0.0003208	0.01791	-0.89
story	(Intercept)	0.008403	0.09167	
	cond	0.0036623	0.06052	-0.19
Residual		0.1104326	0.33231	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.987081	0.220416	0.55506564	1.41909636	36.984369	4.478	0.0000698
lang	-0.012631	0.030082	-0.0715917	0.04632972	35.050283	-0.42	0.6771
cond	-0.034321	0.0348	-0.102529	0.033887	76.932862	-0.986	0.3271
CS_freq	0.016193	0.03365	-0.049761	0.082147	35.704016	0.481	0.6333
Man	-0.130546	0.103728	-0.3338529	0.07276088	35.341735	-1.259	0.2164
Fr	0.054889	0.079058	-0.1000647	0.20984268	36.139664	0.694	0.4919
age	-0.00842	0.005497	-0.0191941	0.00235412	35.164698	-1.532	0.1345
lang:cond	0.020012	0.029164	-0.0371494	0.07717344	791.075583	0.686	0.4928
lang:CS_freq	0.003148	0.008363	-0.0132435	0.01953948	35.258031	0.376	0.7089
cond:CS_freq	0.016629	0.008388	0.00018852	0.03306948	789.786376	1.982	0.0478
lang:cond:CS_freq	-0.004594	0.008109	-0.0204876	0.01129964	787.838522	-0.567	0.5712

RT simplified model (bilinguals)

Random effects:

Groups	Name	Variance	Std.Dev.
runid	(Intercept)	0.079523	0.282
story	(Intercept)	0.008396	0.09163
Residual		0.113604	0.33705

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.716391	0.122642	0.47601268	0.95676932	41.224205	5.841	7.14E-07
cond	-0.032478	0.030497	-0.0922521	0.02729612	840.874978	-1.065	0.2872
CS_freq	0.016734	0.03325	-0.048436	0.081904	37.957693	0.503	0.6177
cond:CS_freq	0.017023	0.008497	0.00036888	0.03367712	840.945807	2.004	0.0454

RT full model (monolinguals)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.0601431	0.24524		
	lang	0.0005581	0.02362	-0.49	
story	(Intercept)	0.0124861	0.11174		
	lang	0.0011543	0.03397	-0.03	
	cond	0.0056111	0.07491	-0.1	0.3
Residual		0.1133204	0.33663		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.895142	0.201077	0.50103108	1.28925292	37.94794	4.452	0.0000725
lang	0.028655	0.027581	-0.0254038	0.08271376	36.735897	1.039	0.306
cond	0.001145	0.03322	-0.0639662	0.0662562	35.14296	0.034	0.973
CS_freq	0.00169	0.061508	-0.1188657	0.12224568	36.11138	0.027	0.978
Man	-0.046839	0.105649	-0.253911	0.16023304	34.962929	-0.443	0.66
Fr	0.098052	0.066443	-0.0321763	0.22828028	34.986328	1.476	0.149
age	-0.006193	0.005729	-0.0174218	0.00503584	35.192155	-1.081	0.287
lang:cond	-0.004848	0.024783	-0.0534227	0.04372668	775.052838	-0.196	0.845
lang:CS_freq	-0.0164	0.017567	-0.0508313	0.01803132	37.864056	-0.934	0.356
cond:CS_freq	-0.008035	0.017335	-0.0420116	0.0259416	800.53722	-0.464	0.643
lang:cond:CS_freq	0.003763	0.016971	-0.0295002	0.03702616	731.88273	0.222	0.825

<u>Appendix III: Experiment 2 Stimuli</u>

Audio stimuli are available on OSF at: https://osf.io/479ry/

Participants were assigned to one of four lists.

All participants heard all six stories (Chicken, Donkey part 1, Donkey part 2, Khalai, Nozibele part 1, Nozibele part 2) twice (once in the French-Mandarin order and once in the Mandarin-French order).

Lists 1 and 3: in the Mandarin-French order of each story Mandarin was from the unnatural condition and French was from the naturalistic condition (anticipatory cues for Mandarin, perseverative cues for French); in the French-Mandarin order of each story French was from the unnatural condition and Mandarin was from the naturalistic condition (anticipatory cues for French, perseverative cues for Mandarin)
Lists 2 and 4: in the Mandarin-French order of each story Mandarin was from the naturalistic condition and French was from the unnatural condition (anticipatory cues for French, perseverative cues for Mandarin); in the French-Mandarin order of each story French was from the naturalistic condition and Mandarin was from the unnatural condition (anticipatory cues for Mandarin, perseverative cues for French)

In lists 1 and 2, Mandarin was on the left and French was on the right.
In lists 3 and 4, French was on the left and Mandarin was on the right.

Appendix IV: Experiment 2 Model Specifications

RT full model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0497696	0.22309	
story	(Intercept)	0.0107294	0.10358	
	cond	0.0006392	0.02528	0.36
Residual		0.0980829	0.31318	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.734159	0.047413	0.64122952	0.82708848	35.039467	15.484	<2.00E-16
lang	-0.002896	0.010362	-0.0232055	0.01741352	847.940888	-0.279	0.77994
cond	-0.02973	0.012675	-0.054573	-0.004887	11.28162	-2.345	0.03826
CSCent	0.042589	0.027917	-0.0121283	0.09730632	34.874057	1.526	0.13613
ManCent	-0.107799	0.077609	-0.2599126	0.04431464	34.936151	-1.389	0.17362
FrCent	0.020234	0.063791	-0.1047964	0.14526436	34.852819	0.317	0.75299
agecent	0.012432	0.004494	0.00362376	0.02124024	34.747625	2.766	0.00902
lang:cond	-0.021066	0.010378	-0.0414069	-0.0007251	846.719629	-2.03	0.04268
lang:CSCent	0.010582	0.007738	-0.0045845	0.02574848	848.216083	1.368	0.1718
cond:CSCent	-0.00369	0.007736	-0.0188526	0.01147256	846.503889	-0.477	0.63346
lang:cond:CSCent	-0.002076	0.007749	-0.017264	0.01311204	846.688258	-0.268	0.78887

RT simplified model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0513477	0.2266	
story	(Intercept)	0.0107392	0.10363	
	cond	0.0006347	0.02519	0.36
Residual		0.0979901	0.31303	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.734199	0.047829	0.64045416	0.82794384	36.615057	15.351	<2e-16
lang	-0.002878	0.010357	-0.0231777	0.01742172	850.954537	-0.278	0.7812
cond	-0.029782	0.012657	-0.0545897	-0.0049743	11.283767	-2.353	0.0377
agecent	0.012168	0.004488	0.00337152	0.02096448	37.778535	2.711	0.01
lang:cond	-0.021049	0.010373	-0.0413801	-0.0007179	849.724476	-2.029	0.0428

ACC full model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.089133	0.29855		
	lang	0.00562	0.07497	0.35	
story	(Intercept)	0.075541	0.27485		
	cond	0.003578	0.05981	-0.09	
	lang	0.005323	0.07296	0.39	-0.81
Residual		0.201928	0.44936		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.915323	0.203351	1.51675504	2.31389096	43.423779	9.419	4.67E-12
lang	-0.063554	0.028395	-0.1192082	-0.0078998	14.629464	-2.238	0.04122
cond	0.058236	0.022795	0.0135578	0.1029142	11.002995	2.555	0.02676
CSCent	-0.058875	0.037559	-0.1324906	0.01474064	34.738499	-1.568	0.12606
ManCent	0.122007	0.102255	-0.0784128	0.3224268	34.805241	1.193	0.24088
FrCent	0.028233	0.08402	-0.1364462	0.1929122	34.674215	0.336	0.73888
age	-0.017855	0.005918	-0.0294543	-0.0062557	34.553637	-3.017	0.00477
lang:cond	0.001798	0.014892	-0.0273903	0.03098632	797.59362	0.121	0.90394
lang:CSCent	-0.018926	0.014223	-0.0468031	0.00895108	36.664325	-1.331	0.19153
cond:CSCent	0.002628	0.011108	-0.0191437	0.02439968	801.147498	0.237	0.81301
lang:cond:CSCent	0.01015	0.01112	-0.0116452	0.0319452	797.505684	0.913	0.36163

ACC simplified model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
runid	(Intercept)	0.089345	0.29891		
	lang	0.005931	0.07701	0.35	
story	(Intercept)	0.075592	0.27494		
	cond	0.003547	0.05955	-0.08	
	lang	0.005329	0.073	0.39	-0.8
Residual		0.201386	0.44876		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.88953	0.20099	1.4955896	2.2834704	46.99996	9.401	2.26E-12
lang	-0.06369	0.02853	-0.1196088	-0.0077712	14.92631	-2.232	0.04135
cond	0.05827	0.02272	0.0137388	0.1028012	11.00443	2.564	0.02632
age	-0.01701	0.00583	-0.0284368	-0.0055832	37.62251	-2.917	0.00592

French – RT full model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0547589	0.23401	
story	(Intercept)	0.0086591	0.09305	
	cond	0.0006014	0.02452	-0.53
Residual		0.0931129	0.30514	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.731313	0.047936	0.63735844	0.82526756	35.17714	15.256	<2e-16
cond	-0.050906	0.015956	-0.0821798	-0.0196322	9.645617	-3.19	0.0101
CSCent	0.050559	0.030126	-0.008488	0.10960596	34.625526	1.678	0.1023
ManCent	-0.081592	0.083783	-0.2458067	0.08262268	34.744224	-0.974	0.3369
FrCent	0.033774	0.068833	-0.1011387	0.16868668	34.594901	0.491	0.6268
agecent	0.009887	0.004849	0.00038296	0.01939104	34.486605	2.039	0.0492
cond:CSCent	-0.005498	0.010676	-0.026423	0.01542696	394.755197	-0.515	0.6069

French – RT simplified model

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0563354	0.23735	
story	(Intercept)	0.0086717	0.09312	
	cond	0.0005969	0.02443	-0.54
Residual		0.0929464	0.30487	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.731765	0.048346	0.63700684	0.82652316	36.972938	15.136	<2e-16
cond	-0.050897	0.015933	-0.0821257	-0.0196683	9.574499	-3.194	0.0101
agecent	0.009696	0.004833	0.00022332	0.01916868	37.525297	2.006	0.0521

Mandarin – RT full model

Random effects:

Groups	Name	Variance	Std.Dev.
runid	(Intercept)	0.04347	0.2085
story	(Intercept)	0.01364	0.1168
Residual		0.10323	0.3213

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.734785	0.049528	0.63771012	0.83185988	27.198152	14.836	1.47E-14
cond	-0.007924	0.015055	-0.0374318	0.0215838	403.972052	-0.526	0.599
CSCent	0.034313	0.02753	-0.0196458	0.0882718	34.762073	1.246	0.221
ManCent	-0.13408	0.076573	-0.2841631	0.01600308	34.899583	-1.751	0.0887
FrCent	0.006383	0.062896	-0.1168932	0.12965916	34.720813	0.101	0.9197
agecent	0.015067	0.004427	0.00639008	0.02374392	34.507655	3.403	0.0017
cond:CSCent	-0.000993	0.011241	-0.0230254	0.02103936	403.92157	-0.088	0.9297

Mandarin – RT simplified model

Random effects:

Groups	Name	Variance	Std.Dev.
runid	(Intercept)	0.0429	0.2071
story	(Intercept)	0.01364	0.1168
Residual		0.10281	0.3206

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.735168	0.049372	0.63839888	0.83193712	27.366759	14.89	1.20E-14
agecent	0.014756	0.004334	0.00626136	0.02325064	36.507717	3.405	0.00162
ManCent	-0.124301	0.075673	-0.2726201	0.02401808	36.817451	-1.643	0.10898

Appendix V: Experiment 3 Stimuli

The same audio stimuli from experiment 1 were used in experiment 3.

Participants were assigned to one of eight lists.

All participants heard all six stories (Chicken, Donkey part 1, Donkey part 2, Khalai, Nozibele part 1, Nozibele part 2) once (three in the French-Mandarin order and three in the Mandarin-French order).

Chicken, Donkey part 1, and Nozibele part 1 were considered "group A stories" and were always in one condition.
Donkey part 2, Khalai, and Nozibele part 2 were considered "group B stories" and were always in the other condition.

In lists 1, 3, 5, and 7, group A stories were in the French-Mandarin order and group B stories were in the Mandarin-French order
In lists 2, 4, 6, and 8, group A stories were in the Mandarin-French order and group B stories were in the French-Mandarin order

In lists 1, 4, 5, and 8, the French-Mandarin stories were presented in the naturalistic condition and the Mandarin-French stories were presented in the unnatural condition.
In lists 2, 3, 6, and 7, the French-Mandarin stories were presented in the unnatural condition and the Mandarin-French stories were presented in the naturalistic condition.

In lists 1, 2, 3, and 4, Mandarin was on the left and French was on the right.
In lists 5, 6, 7, and 8, French was on the left and Mandarin was on the right.

Appendix VI: Experiment 3 Model Specifications

Hit time

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.001769	0.04206	
	lang	0.002288	0.04783	-0.14
Residual		0.022248	0.14916	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.465863	0.009697	0.44685688	0.48486912	38.605587	48.042	<2e-16
lang	-0.012803	0.010347	-0.0330831	0.00747712	35.266397	-1.237	0.224
cond	-0.004161	0.00706	-0.0179986	0.0096766	371.216873	-0.589	0.556
ageyrscent	0.011979	0.00927	-0.0061902	0.0301482	37.968933	1.292	0.204
lang:cond	-0.008043	0.007064	-0.0218884	0.00580244	371.232941	-1.139	0.256
lang:ageyrscent	-0.007529	0.009897	-0.0269271	0.01186912	34.733652	-0.761	0.452
cond:ageyrscent	-0.004411	0.006717	-0.0175763	0.00875432	368.786523	-0.657	0.512
lang:cond:ageyrscent	-0.005885	0.00672	-0.0190562	0.0072862	368.771537	-0.876	0.382

Correct Rejection time

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.001523	0.03902	
	lang	0.002445	0.04945	0.2
Residual		0.02265	0.1505	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.5493	0.00942	0.5308368	0.5677632	37.92	58.312	<2e-16
lang	0.01043	0.01058	-0.0103068	0.0311668	35.7	0.986	0.331
cond	0.003722	0.00712	-0.0102332	0.0176772	371.6	0.523	0.601
ageyrscent	0.002567	0.009003	-0.0150789	0.02021288	37.26	0.285	0.777
lang:cond	0.009256	0.007129	-0.0047168	0.02322884	370.9	1.298	0.195
lang:ageyrscent	0.004822	0.01012	-0.0150132	0.0246572	35.17	0.477	0.637
cond:ageyrscent	-0.0002177	0.006776	-0.0134987	0.01306326	369	-0.032	0.974
lang:cond:ageyrscent	0.0006533	0.006781	-0.0126375	0.01394406	368.5	0.096	0.923

Correct Moves

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
runid	(Intercept)	0.0009592	0.03097	
	lang	0.0009733	0.0312	0.75
Residual		0.0357272	0.18902	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.5166	0.01018	0.4966472	0.5365528	38.85	50.745	<2e-16
lang	-0.02845	0.0102	-0.048442	-0.008458	36.65	-2.79	0.00832
cond	-0.01256	0.008926	-0.030055	0.00493496	375.8	-1.408	0.16009
ageyrscent	0.0007214	0.009715	-0.01832	0.0197628	38.02	0.074	0.94119
lang:cond	0.005596	0.008926	-0.011899	0.02309096	376.3	0.627	0.53113
lang:ageyrscent	0.006856	0.009732	-0.0122187	0.02593072	35.87	0.704	0.48567
cond:ageyrscent	-0.007772	0.0085	-0.024432	0.008888	372.8	-0.914	0.3611
lang:cond:ageyrscent	0.01212	0.0085	-0.00454	0.02878	373.2	1.426	0.15458

Appendix VII: Experiment 4 Stimuli

All stimuli are available on OSF at: https://osf.io/479ry/

Participants were assigned to one of four lists.

In lists 1 and 3, all stimuli were derived from the unnatural condition
In lists 2 and 4, all stimuli were derived from the naturalistic condition

In lists 1 and 2, French was on the left and Mandarin was on the right
In lists 3 and 4, Mandarin was on the left and French was on the right

Training phase: trial details

trial	*type*	*category*	*story*	*gap*	*naturalistic: midpoint*	*naturalistic: restart point*	*unnatural: midpoint*	*unnatural: restart point*
1	FM	switch	Nozibele2_ManFr	0	3.695	3.695	3.758	3.758
2	FF	stay	Nozibele2_FrMan	80	3.979	4.059	3.666	3.746
3	MF	switch	Khalai_ManFr	160	2.915	3.075	3.066	3.226
4	MM	stay	Khalai_FrMan	240	2.853	3.093	2.903	3.143
5	MF	switch	Donkey2_ManFr	320	3.634	3.954	4.014	4.334
6	FM	switch	Nozibele2_ManFr	400	4.062	4.462	4.2	4.6
7	FF	stay	Donkey2_FrMan	480	4.767	5.247	4.435	4.915
8	MM	stay	Nozibele2_FrMan	560	3.256	3.816	3.301	3.861
9	FM	switch	Khalai_ManFr	640	2.422	3.062	2.284	2.924
10	FF	stay	Khalai_FrMan	720	3.258	3.978	3.277	3.997
11	MF	switch	Donkey2_ManFr	800	2.932	3.732	3.262	4.062
12	MM	stay	Donkey2_FrMan	880	3.368	4.248	3.451	4.331
13	FF	stay	Nozibele2_FrMan	960	4.361	5.321	3.767	4.727
14	FM	switch	Donkey2_ManFr	1040	3.695	4.735	3.215	4.255
15	MF	switch	Nozibele2_ManFr	1120	4.45	5.57	3.929	5.049
16	FF	stay	Khalai_FrMan	1200	3.517	4.717	3.338	4.538
17	MM	stay	Donkey2_FrMan	1280	3.383	4.663	3.309	4.589
18	MF	switch	Khalai_ManFr	1360	4.254	5.614	4.298	5.658
19	FM	switch	Khalai_ManFr	1440	3.196	4.636	3.152	4.592
20	MM	stay	Nozibele2_FrMan	1520	2.994	4.514	3.329	4.849
21	FF	stay	Donkey2_FrMan	1600	4.548	6.148	3.897	5.497
22	FM	switch	Donkey2_ManFr	1680	3.181	4.861	2.96	4.64
23	MF	switch	Nozibele2_ManFr	1760	3.524	5.284	3.236	4.996
24	MM	stay	Khalai_FrMan	1840	3.068	4.908	3.003	4.843

Test phase: trial details

trial	*type*	*category*	*story*	*naturalistic: midpoint*	*naturalistic: restart point*	*unnatural: midpoint*	*unnatural: restart point*
1	MF	switch	Chicken_ManFr	3.465	5.465	3.667	5.667
2	FF	stay	Chicken_FrMan	4.195	6.195	3.324	5.324
3	MM	stay	Donkey1_FrMan	3.992	5.992	3.209	5.209
4	FM	switch	Donkey1_ManFr	2.707	4.707	2.7	4.7
5	MF	switch	Nozibele1_ManFr	3.148	5.148	3.257	5.257
6	MM	stay	Nozibele1_FrMan	4.321	6.321	4.132	6.132
7	FM	switch	Chicken_ManFr	4.799	6.799	4.291	6.291
8	FF	stay	Donkey1_FrMan	3.793	5.793	3.659	5.659
9	MM	stay	Chicken_FrMan	3.788	5.788	3.895	5.895
10	MF	switch	Donkey1_ManFr	3.491	5.491	3.622	5.622
11	FM	switch	Nozibele1_ManFr	4.41	6.41	4.238	6.238
12	FF	stay	Nozibele1_FrMan	3.741	5.741	3.728	5.728
13	MF	switch	Chicken_ManFr	3.292	5.292	3.418	5.418
14	MM	stay	Donkey1_FrMan	3.528	5.528	3.454	5.454
15	FF	stay	Chicken_FrMan	3.119	5.119	2.959	4.959
16	FM	switch	Donkey1_ManFr	4.092	6.092	3.9	5.9
17	MF	switch	Nozibele1_ManFr	4.337	6.337	3.937	5.937
18	FF	stay	Nozibele1_FrMan	4.056	6.056	3.8	5.8
19	MM	stay	Chicken_FrMan	4.082	6.082	4.397	6.397
20	MF	switch	Donkey1_ManFr	3.566	5.566	3.851	5.851
21	FM	switch	Chicken_ManFr	4.544	6.544	4.259	6.259
22	FF	stay	Donkey1_FrMan	4.051	6.051	4.484	6.484
23	FM	switch	Nozibele1_ManFr	2.832	4.832	2.811	4.811
24	MM	stay	Nozibele1_FrMan	3.952	5.952	3.487	5.487

Appendix VIII: Experiment 4 Complete Data

ID	*Order*	*Condition*	*Age*	*Sex*	*Training # of trials with looks*	*Training attention%*	*Training # of switch trials with looks*	*Training attention % in switch trials*	*Training right accurate look %*	*Training left accurate look %*
4	B	naturalistic	9	m	4	0.26434639	1	0.11300676	1	0.6666667
7	B	naturalistic	11	m	1	0.03090278	0	0	n/a	1
8	A	unnatural	7	f	4	0.49051724	1	0.23103448	0.78571429	0
10	A	unnatural	12	f	8	0.78250946	4	0.80574324	0.66666667	0.8
13	A	unnatural	7	m	6	0.52307299	3	0.55177833	0.71092089	0.51263244
18	A	unnatural	13	f	1	0.125	0	0	n/a	1
22	C	unnatural	9	m	6	0.49961985	2	0.29262821	0.56565657	0.72016533
20	C	unnatural	11	f	5	0.4021515	3	0.41733871	0.91452991	0.66846869
34	D	naturalistic	6	m	5	0.2867258	2	0.16723406	0	0.38107988
31	C	unnatural	9	m	2	0.13389591	0	0	1	0
27	C	unnatural	8	f	8	0.75296092	4	0.78640076	0.51487213	0.62216532
32	D	naturalistic	12	f	4	0.26112374	2	0.30763027	1	0.33333333
33	D	naturalistic	6	f	7	0.46314394	3	0.37089202	0.14375	0.49224008
41	D	naturalistic	12	f	1	0.06818182	0	0	n/a	0
42	D	naturalistic	7	m	3	0.2482585	1	0.25	1	1
28	C	unnatural	14	f	4	0.38036077	1	0.25	0	0.6224359
38	D	naturalistic	11	f	0	0	0	0	n/a	n/a

Test attention%	*Test # of trials with looks*	*Training FM %*	*Training MF %*	*Training FF%*	*Training MM %*	*Training overall accuracy %*	*Training switch accurate look %*	*Training stay accurate look %*
0.09430556	5	0	n/a	1	1	0.75	0	1
0.18763889	7	n/a	n/a	n/a	1	1	n/a	1
0.3474537	11	n/a	1	0.57142857	0	0.39285714	1	0.19047619
0.63583333	15	1	0.5	1	0.66666667	0.75	0.75	0.75
0.46823671	17	0.53789731	0.56638133	1	0.5	0.61177666	0.55688666	0.66666667
0.12627315	6	n/a	n/a	n/a	1	1	n/a	1
0.18939815	10	n/a	0.63448276	0.89153046	0.56565657	0.64291095	0.63448276	0.64712504
0.64298611	18	0.82474227	0.87179487	1	0.51219512	0.81610543	0.85611067	0.75609756
0.12645833	10	0	0	0	0.57161981	0.22864793	0	0.38107988
0.1210101	5	n/a	n/a	1	0	0.5	n/a	0.5
0.53134259	18	0.72405914	0.77230819	0	0.55423611	0.58193037	0.74818366	0.41567708
0.05349206	5	0	1	n/a	0.5	0.5	0.5	0.5
0.33912037	16	0.56743003	0	0.2875	0.44211344	0.39267148	0.37828668	0.40346008
0.17358796	9	n/a	n/a	n/a	0	0	n/a	0
0.36543981	14	n/a	1	n/a	1	1	1	1
0.29076389	14	0	n/a	0	0.93365385	0.46682692	0	0.6224359
0.24386243	11	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Test FF%	*Test MM%*	*Test overall accuracy %*	*Test switch accurate look %*	*Test stay accurate look %*	*Test left accurate look %*	*Test right accurate look %*	*Test attention % in switch trials*	*Test # of switch trials with looks*
n/a	0.06960557	0.39751349	0.61611877	0.06960557	0.32918915	0.5	0.10838889	3
0.25949954	0.27083333	0.32430742	0.40316206	0.26516643	0.51388889	0.18212131	0.16092593	3
0.6292922	0.67722222	0.69387019	0.78291667	0.64298649	0.77611111	0.64687538	0.27222222	4
0.32022222	1	0.50674074	0.42857143	0.57513889	0.85714286	0.20013889	0.59074074	7
0.71446043	0.7474817	0.65452862	0.55004245	0.72766894	0.72101722	0.6079866	0.33646465	7
0.5	1	0.58738024	0.26214073	0.75	0.84142716	0.33333333	0.1300463	2
1	0.62046632	0.55510142	0.41382979	0.69637306	0.4375	0.63350237	0.20888889	5
0.4744795	0.44144737	0.49468084	0.52405476	0.45796343	0.35263828	0.59131382	0.76606481	10
1	0.5	0.65724508	0.59540847	0.75	0.75	0.59540847	0.17467593	6
0.65384615	0.36931865	0.48312965	n/a	0.48312965	0.36931865	0.65384615	0	0
0.48924899	0.50945085	0.58270037	0.66717315	0.4982276	0.65752715	0.46982106	0.47361111	9
1	0.06040268	0.61208054	0	0.76510067	0.03020134	1	0.03626263	1
0	0.85074627	0.52647001	0.55783724	0.48614072	0.92537313	0.12756689	0.42712963	9
0	1	0.36914675	0.26446414	0.5	0.75	0.06446414	0.16277778	5
0.75	0.71428571	0.68367347	0.6	0.73015873	0.79591837	0.57142857	0.26657407	5
0.5	0.51182171	0.39539037	0.28571429	0.50506645	0.58924419	0.25	0.24101852	7
0.66666667	0.83026316	0.42368421	0.16666667	0.73210526	0.53210526	0.33333333	0.18737374	6

Test FM%	*Test MF%*
0.84835631	0.5
1	0.10474308
0.875	0.69083333
0.75	0
0.68573124	0.44827586
0.52428147	0
0.65957447	0.25
0.40132583	0.70814815
1	0.3931127
n/a	n/a
0.84448276	0.44553613
0	n/a
1	0.20410703
0.5	0.10744024
1	0.33333333
0.66666667	0
0.33333333	0

Appendix IX: Experiment 5 Model Specifications

Analysis 1

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
participant	(Intercept)	0.020594	0.14351		
	reg	0.002214	0.04705	-0.33	
	cont	0.001428	0.03779	-0.56	-0.24
Residual		0.027514	0.16587		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	5.428	0.04569	5.3384476	5.5175524	8.903	118.814	1.47E-15
reg	-0.126	0.01576	-0.1568896	-0.0951104	9.103	-7.998	0.0000207
cont	0.009926	0.01299	-0.0155344	0.0353864	8.95	0.764	0.46462
reg:cont	-0.0152	0.004861	-0.0247276	-0.0056724	1205	-3.127	0.00181

Analysis 1 – Split by register

IDS only:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
participant	(Intercept)	0.028326	0.1683	
	cont	0.002893	0.05379	0.01
Residual		0.029017	0.17034	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	5.5538	0.05372	5.4485088	5.6590912	8.90415	103.386	5.06E-15
cont	0.0257	0.01848	-0.0105208	0.0619208	9.47627	1.391	0.196

ADS only:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
participant	(Intercept)	0.018344	0.13544	
	cont	0.002822	0.05313	-0.49
Residual		0.022761	0.15087	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	5.297561	0.043512	5.21227748	5.38284452	8.898943	121.751	1.20E-15
cont	-0.006347	0.018328	-0.0422699	0.02957588	8.399038	-0.346	0.738

Analysis 2 – VOT as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
Participant	(Intercept)	0.131036	0.36199		
	reg	0.002638	0.05136	-0.59	
	type	0.011055	0.10514	-0.43	0.27
Phoneme	(Intercept)	0.051174	0.22622		
	reg	0.00177	0.04207	-0.06	
Residual		0.248173	0.49817		

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.479	0.2153	2.057012	2.900988	12.79	11.512	4.05E-08
reg	0.04813	0.03205	-0.014688	0.110948	3.74	1.502	0.2124
type	-0.002379	0.03601	-0.0729586	0.0682006	7.334	-0.066	0.9491
log(WordLength*1000)	0.2309	0.0214	0.188956	0.272844	1923	10.795	<2e-16
reg:type	0.02523	0.01273	0.0002792	0.0501808	1710	1.982	0.0477

Analysis 2 – consonant-vowel ratio as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
Participant	(Intercept)	0.142661	0.3777		
	reg	0.007313	0.08551	0.33	
	type	0.017752	0.13324	-0.63	-0.35
Phoneme	(Intercept)	0.033929	0.1842		
	reg	0.000633	0.02516	0.54	
Residual		0.510352	0.71439		

Fixed effects:

<table>
<tr><th></th><th>Estimate</th><th>Std.Error</th><th>95%CI-low</th><th>95%CI-high</th><th>df</th><th>t value</th><th>Pr(>|t|)</th></tr>
<tr><td>(Intercept)</td><td>0.19413</td><td>0.24228</td><td>-0.2807388</td><td>0.6689988</td><td>37.3968</td><td>0.801</td><td>0.4281</td></tr>
<tr><td>reg</td><td>0.11601</td><td>0.03618</td><td>0.0450972</td><td>0.1869228</td><td>6.79072</td><td>3.206</td><td>0.0156</td></tr>
<tr><td>type</td><td>-0.04265</td><td>0.0464</td><td>-0.133594</td><td>0.048294</td><td>7.70201</td><td>-0.919</td><td>0.3859</td></tr>
<tr><td>log(WordLength*1000</td><td>-0.14681</td><td>0.03061</td><td>-0.2068056</td><td>-0.0868144</td><td>1895.95622</td><td>-4.797</td><td>1.74E-06</td></tr>
<tr><td>reg:type</td><td>0.03437</td><td>0.01811</td><td>-0.0011256</td><td>0.0698656</td><td>1213.58948</td><td>1.898</td><td>0.058</td></tr>
</table>

Analysis 2 – split by register (VOT)

IDS:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.14321	0.3784	
	type	0.03349	0.183	-0.32
Phoneme	(Intercept)	0.05045	0.2246	
Residual		0.27926	0.5284	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	3.79707	0.17766	3.4488564	4.1452836	5.84642	21.372	8.97E-07
type	-0.02879	0.06142	-0.1491732	0.0915932	8.18413	-0.469	0.651
WordLength_cent	0.44207	0.07403	0.2969712	0.5871688	968.89488	5.972	3.29E-09

ADS:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.143346	0.37861	
	type	0.003822	0.06183	0.17
Phoneme	(Intercept)	0.049455	0.22239	
Residual		0.211634	0.46004	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	3.87665	0.17671	3.5302984	4.2230016	5.87533	21.937	7.32E-07
type	0.02299	0.02683	-0.0295968	0.0755768	7.44272	0.857	0.418
WordLength_cent	0.63319	0.07911	0.4781344	0.7882456	944.85994	8.004	3.52E-15

Analysis 2 – split by register (consonant-vowel ratio)

IDS:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.12944	0.3598	
	type	0.03737	0.1933	-0.21
Phoneme	(Intercept)	0.02478	0.1574	
Residual		0.53197	0.7294	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	-0.76809	0.14836	-1.0588756	-0.4773044	8.08738	-5.177	0.000817
type	-0.07543	0.06732	-0.2073772	0.0565172	8.36457	-1.12	0.293681
WordLength_cent	-0.27234	0.10199	-0.4722404	-0.0724396	973.4706	-2.67	0.007705

ADS:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.18604	0.4313	
	type	0.01504	0.1226	-0.44
Phoneme	(Intercept)	0.03582	0.1893	
Residual		0.48168	0.694	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	-0.556684	0.177272	-0.9041371	-0.2092309	8.177205	-3.14	0.0134
type	-0.003755	0.047699	-0.097245	0.08973504	8.777154	-0.079	0.939
WordLength_cent	-0.215746	0.119132	-0.4492447	0.01775272	945.835573	-1.811	0.0705

Analysis 2 – split by story type (VOT)

Unilingual:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.173215	0.41619	
	reg	0.008305	0.09113	-0.66
Phoneme	(Intercept)	0.051327	0.22655	
Residual		0.224607	0.47393	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.562	0.2358	2.099832	3.024168	16.76	10.864	5.29E-09
reg	0.01174	0.03234	-0.0516464	0.0751264	9.033	0.363	0.725
log(WordLength*1000	0.217	0.02463	0.1687252	0.2652748	1262	8.81	<2e-16

Code-switching:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.11508	0.3392	
	reg	0.01224	0.1106	-0.04
Phoneme	(Intercept)	0.03395	0.1843	
Residual		0.28452	0.5334	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.37605	0.28897	1.8096688	2.9424312	73.21431	8.223	5.27E-12
reg	0.08167	0.04345	-0.003492	0.166832	3.5809	1.88	0.142
log(WordLength*1000)	0.24814	0.04117	0.1674468	0.3288332	649.48967	6.027	2.80E-09

Analysis 2 – split by story type (consonant-vowel ratio)

Unilingual:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.22429	0.4736	
	reg	0.01048	0.1024	0.33
Phoneme	(Intercept)	0.01578	0.1256	
Residual		0.46083	0.6788	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.24584	0.26641	-0.2763236	0.7680036	65.29144	0.923	0.3595
reg	0.08028	0.03861	0.0046044	0.1559556	9.04551	2.08	0.0672
log(WordLength*1000)	-0.14767	0.03522	-0.2167012	-0.0786388	1263.90973	-4.193	2.94E-05

Code-switching:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.09609	0.31	
	reg	0.02536	0.1593	0.03
Phoneme	(Intercept)	0.0815	0.2855	
Residual		0.58041	0.7618	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.28007	0.39961	-0.5031656	1.0633056	57.80472	0.701	0.4862
reg	0.10805	0.06228	-0.0140188	0.2301188	7.4676	1.735	0.12366
log(WordLength*1000)	-0.17044	0.05864	-0.2853744	-0.0555056	656.93734	-2.907	0.00378

Analysis 3 – VOT as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.11505	0.3392	
	reg	0.01255	0.112	-0.08
Phoneme	(Intercept)	0.03359	0.1833	
Residual		0.28555	0.5344	

Fixed effects:

<table>
<tr><th></th><th>Estimate</th><th>Std.Error</th><th>95%CI-low</th><th>95%CI-high</th><th>df</th><th>t value</th><th>Pr(>|t|)</th></tr>
<tr><td>(Intercept)</td><td>2.3878</td><td>0.29244</td><td>1.8146176</td><td>2.9609824</td><td>77.63577</td><td>8.165</td><td>4.64E-12</td></tr>
<tr><td>switch2</td><td>-0.28555</td><td>0.23801</td><td>-0.7520496</td><td>0.1809496</td><td>631.53916</td><td>-1.2</td><td>0.231</td></tr>
<tr><td>reg</td><td>-0.04434</td><td>0.24668</td><td>-0.5278328</td><td>0.4391528</td><td>465.62045</td><td>-0.18</td><td>0.857</td></tr>
<tr><td>log(WordLength*1000)</td><td>0.24652</td><td>0.04187</td><td>0.1644548</td><td>0.3285852</td><td>644.51908</td><td>5.888</td><td>6.29E-09</td></tr>
<tr><td>switch2:reg</td><td>-0.03188</td><td>0.23616</td><td>-0.4947536</td><td>0.4309936</td><td>646.62617</td><td>-0.135</td><td>0.893</td></tr>
<tr><td>switch2:log(WordLength*1000)</td><td>0.04932</td><td>0.04008</td><td>-0.0292368</td><td>0.1278768</td><td>632.3367</td><td>1.231</td><td>0.219</td></tr>
<tr><td>reg:log(WordLength*1000)</td><td>0.0215</td><td>0.04098</td><td>-0.0588208</td><td>0.1018208</td><td>628.92405</td><td>0.525</td><td>0.6</td></tr>
<tr><td>switch2:reg:log(WordLength*1000)</td><td>0.00762</td><td>0.03976</td><td>-0.0703096</td><td>0.0855496</td><td>646.81066</td><td>0.192</td><td>0.848</td></tr>
</table>

Analysis 3 – consonant-vowel ratio as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.09404	0.3067	
	reg	0.02383	0.1544	-0.03
Phoneme	(Intercept)	0.08175	0.2859	
Residual		0.58197	0.7629	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.32092	0.40458	-0.4720568	1.1138968	60.23739	0.793	0.43076
switch2	-0.4002	0.33929	-1.0652084	0.2648084	642.42709	-1.18	0.23863
reg	-0.00529	0.35145	-0.694132	0.683552	544.80124	-0.015	0.988
log(WordLength*1000)	-0.17677	0.05958	-0.2935468	-0.0599932	651.87441	-2.967	0.00312
switch2:reg	-0.23139	0.33698	-0.8918708	0.4290908	647.44636	-0.687	0.49255
switch2:log(WordLength*1000)	0.06666	0.05713	-0.0453148	0.1786348	642.90025	1.167	0.24375
reg:log(WordLength*1000)	0.01973	0.05843	-0.0947928	0.1342528	638.80797	0.338	0.73573
switch2:reg:log(WordLength*1000)	0.03978	0.05673	-0.0714108	0.1509708	647.69149	0.701	0.48343

Analysis 4 – VOT as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.12345	0.3514	
	reg	0.01143	0.1069	-0.2
Phoneme	(Intercept)	0.032	0.1789	
Residual		0.26037	0.5103	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.80353	0.37797	2.0627088	3.5443512	145.75571	7.417	9.07E-12
RelevantDistance	0.02759	0.03769	-0.0462824	0.1014624	329.40478	0.732	0.46467
switchdirection	-0.07127	0.05693	-0.1828528	0.0403128	324.53611	-1.252	0.21149
reg	0.01094	0.06851	-0.1233396	0.1452196	15.80085	0.16	0.87512
log(WordLength*1000)	0.16968	0.05721	0.0575484	0.2818116	335.29768	2.966	0.00323
RelevantDistance:switch direction	0.04615	0.03674	-0.0258604	0.1181604	325.26268	1.256	0.20992
RelevantDistance:reg	0.041	0.03733	-0.0321668	0.1141668	337.34566	1.098	0.2729
switchdirection:reg	-0.05955	0.05681	-0.1708976	0.0517976	330.34071	-1.048	0.29534
RelevantDistance:switch direction:reg	0.02395	0.03684	-0.0482564	0.0961564	329.69878	0.65	0.51606

Analysis 4 – consonant-vowel ratio as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.12112	0.348	
	reg	0.02632	0.1622	-0.13
Phoneme	(Intercept)	0.07152	0.2674	
Residual		0.53243	0.7297	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.96735	0.52895	-0.069392	2.004092	140.13563	1.829	0.069553
RelevantDistance	-0.01722	0.05376	-0.1225896	0.0881496	335.83738	-0.32	0.748885
switchdirection	-0.17212	0.08134	-0.3315464	-0.0126936	330.37373	-2.116	0.035079
reg	0.03579	0.09964	-0.1595044	0.2310844	29.00464	0.359	0.722054
log(WordLength*1000)	-0.27753	0.08166	-0.4375836	-0.1174764	339.44209	-3.398	0.000758
RelevantDistance:switch direction	0.09217	0.05248	-0.0106908	0.1950308	331.16117	1.756	0.079969
RelevantDistance:reg	0.06042	0.05336	-0.0441656	0.1650056	338.53243	1.132	0.258273
switchdirection:reg	-0.07869	0.08125	-0.23794	0.08056	332.36124	-0.968	0.333506
RelevantDistance:switch direction:reg	0.03199	0.05269	-0.0712824	0.1352624	332.31215	0.607	0.544192

Analysis 4 –adult-directed anticipatory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1254	0.3541
Phoneme	(Intercept)	0.1904	0.4364
Residual		0.1589	0.3987

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.49361	0.63824	1.2426596	3.7445604	44.69775	3.907	0.000313
RelevantDistance	0.14375	0.06392	0.0184668	0.2690332	77.82031	2.249	0.027353
log(WordLength*1000)	0.19004	0.0945	0.00482	0.37526	79.38921	2.011	0.047708

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.2523	0.5023
Phoneme	(Intercept)	0.1579	0.3974
Residual		0.3681	0.6067

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.10776	0.9169	-1.689364	1.904884	77.25874	0.118	0.9068
RelevantDistance	0.2032	0.09712	0.0128448	0.3935552	78.88448	2.092	0.0396
log(WordLength*1000)	-0.19021	0.14343	-0.4713328	0.0909128	80.70807	-1.326	0.1885

Fixed consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1332	0.365
Phoneme	(Intercept)	0.1491	0.3861
Residual		0.1438	0.3793

Fixed effects:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	2.09418	0.61773	53.0995	3.39	0.00132
RelevantDistance	0.11795	0.06145	75.83417	1.919	0.05868
log(WordLength*1000)	0.0824	0.09649	76.43012	0.854	0.39582
log(VowelLength_usethisone*1000)	-0.76777	0.07591	83.00826	-10.114	3.91E-16

Analysis 4 –adult-directed perseveratory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1124	0.3353
Phoneme	(Intercept)	0.1738	0.4169
Residual		0.2383	0.4881

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.9965	0.75974	0.5074096	3.4855904	59.97628	2.628	0.01089
RelevantDistance	-0.02957	0.06965	-0.166084	0.106944	78.32226	-0.425	0.67232
log(WordLength*1000)	0.32088	0.11987	0.0859348	0.5558252	78.88446	2.677	0.00903

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.03453	0.1858
Phoneme	(Intercept)	0.1373	0.3705
Residual		0.57246	0.7566

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.9352	1.0697	-1.161412	3.031812	73.8702	0.874	0.385
RelevantDistance	-0.1276	0.1036	-0.330656	0.075456	81.6018	-1.232	0.222
log(WordLength*1000)	-0.2223	0.1758	-0.566868	0.122268	78.8055	-1.264	0.21

Fixed consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1079	0.3285
Phoneme	(Intercept)	0.1587	0.3984
Residual		0.2377	0.4875

Fixed effects:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1.81407	0.76866	63.81827	2.36	0.0213
RelevantDistance	-0.02723	0.06954	77.31266	-0.392	0.6965
log(WordLength*1000)	0.2578	0.12926	75.613	1.994	4.97E-02
log(VowelLength_usethisone*1000)	-0.8768	0.09651	74.0916	-9.085	1.14E-13

Analysis 4 –infant-directed anticipatory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1534	0.3917
Phoneme	(Intercept)	0.02066	0.1437
Residual		0.26873	0.5184

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	3.8142581	0.7604735	2.32373004	5.30478616	95.0030728	5.016	2.45E-06
RelevantDistance	-0.0001773	0.0678117	-0.1330882	0.13273363	93.905422	-0.003	0.998
log(WordLength*1000)	-0.0003161	0.1247996	-0.2449233	0.24429112	92.5270676	-0.003	0.998

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1721	0.4149
Phoneme	(Intercept)	0.1089	0.33
Residual		0.4969	0.7049

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.25788	1.04281	0.2139724	4.3017876	92.83625	2.165	0.03294
RelevantDistance	-0.04037	0.09216	-0.2210036	0.1402636	94.20939	-0.438	0.66233
log(WordLength*1000)	-0.512	0.17051	-0.8461996	-0.1778004	93.653	-3.003	0.00343

Fixed consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.14654	0.3828
Phoneme	(Intercept)	0.01895	0.1377
Residual		0.26988	0.5195

Fixed effects:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	3.767778	0.76093	94.015009	4.952	3.23E-06
RelevantDistance	0.001398	0.067925	92.895104	0.021	0.984
log(WordLength*1000)	-0.103243	0.161266	88.171934	-0.64	0.524
log(VowelLength_usethisone*1000)	-0.866642	0.128752	90.644285	-6.731	1.49E-09

Analysis 4 –infant-directed perseveratory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.200996	0.44833
Phoneme	(Intercept)	0.002114	0.04598
Residual		0.307491	0.55452

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.7228	0.6838	1.382552	4.063048	70.2626	3.982	0.000165
RelevantDistance	-0.0605	0.0944	-0.245524	0.124524	70.363	-0.641	0.523641
log(WordLength*1000)	0.1878	0.1076	-0.023096	0.398696	68.2445	1.746	0.085249

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.298789	0.5466
Phoneme	(Intercept)	0.004679	0.0684
Residual		0.602546	0.7762

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.1848	0.9467	-1.670732	2.040332	70.1388	0.195	0.846
RelevantDistance	-0.1519	0.1313	-0.409248	0.105448	70.8245	-1.157	0.251
log(WordLength*1000)	-0.1295	0.15	-0.4235	0.1645	68.8146	-0.863	0.391

Fixed consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.2275	0.477
Residual		0.463	0.6804

Fixed effects:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	2.917	0.9234	70.8866	3.159	0.00233
RelevantDistance	-0.1933	0.1159	69.8791	-1.668	0.09976
log(WordLength*1000)	-0.2081	0.1418	67.6432	-1.467	0.14692
log(VowelLength_usethisone*1000)	-0.5426	0.1667	70.2537	-3.256	0.00174

Analysis 4b – VOT as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.13212	0.36348	
	reg	0.00733	0.08562	-0.31
Phoneme	(Intercept)	0.02652	0.16285	
Residual		0.2614	0.51128	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.369473	0.316603	1.74893112	2.99001488	119.008769	7.484	1.37E-11
RelevantDistance_using5	0.013964	0.01817	-0.0216492	0.0495772	472.017005	0.768	0.443
switchdirection_using5	-0.035222	0.043719	-0.1209112	0.05046724	468.051965	-0.806	0.421
reg	0.052184	0.052771	-0.0512472	0.15561516	11.477262	0.989	0.343
log(WordLength*1000)	0.24918	0.046502	0.15803608	0.34032392	476.049609	5.358	1.31E-07
RelevantDistance_using5:switchdirection_using5	0.011098	0.017801	-0.023792	0.04598796	470.148689	0.623	0.533
RelevantDistance_using5:reg	0.012701	0.018132	-0.0228377	0.04823972	474.707346	0.7	0.484
switchdirection_using5:reg	-0.029823	0.043532	-0.1151457	0.05549972	468.657669	-0.685	0.494
RelevantDistance_using5:switchdirection_using5:reg	0.005171	0.017756	-0.0296308	0.03997276	462.648298	0.291	0.771

Analysis 4b –Consonant-vowel ratio as dependent variable

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.14419	0.3797	
	reg	0.02423	0.1557	0.21
Phoneme	(Intercept)	0.08122	0.285	
Residual		0.49818	0.7058	

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.413212	0.43705	-0.443406	1.26983	75.438867	0.945	0.34744
RelevantDistance_using5	0.007135	0.025082	-0.0420257	0.05629572	471.999514	0.284	0.77619
switchdirection_using5	-0.146136	0.060366	-0.2644534	-0.0278186	468.80647	-2.421	0.01586
reg	0.083673	0.080427	-0.0739639	0.24130992	16.365519	1.04	0.3133
log(WordLength*1000)	-0.18787	0.064312	-0.3139215	-0.0618185	475.014658	-2.921	0.00365
RelevantDistance_using5:switchdirection_using5	0.052862	0.0246	0.004646	0.101078	469.121192	2.149	0.03215
RelevantDistance_using5:reg	0.018022	0.025052	-0.0310799	0.06712392	473.907905	0.719	0.47225
switchdirection_using5:reg	-0.057694	0.060135	-0.1755586	0.0601706	469.108561	-0.959	0.33785
RelevantDistance_using5:switchdirection_using5:reg	0.021727	0.024517	-0.0263263	0.06978032	466.813138	0.886	0.37597

Analysis 4b –adult-directed anticipatory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1073	0.3276
Phoneme	(Intercept)	0.1063	0.326
Residual		0.2047	0.4524

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.49549	0.53034	1.4560236	3.5349564	62.70128	4.705	1.44E-05
RelevantDistance_using5	0.0373	0.02887	-0.0192852	0.0938852	122.66908	1.292	0.19878
log(WordLength*1000)	0.21432	0.08072	0.0561088	0.3725312	123.82746	2.655	0.00897

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1951	0.4417
Phoneme	(Intercept)	0.135	0.3675
Residual		0.3992	0.6318

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.59758	0.72223	-0.8179908	2.0131508	89.4519	0.827	0.4102
RelevantDistance_using5	0.09501	0.0403	0.016022	0.173998	123.36975	2.357	0.02
log(WordLength*1000)	-0.24957	0.11257	-0.4702072	-0.0289328	124.775	-2.217	0.0284

Analysis 4b –adult-directed perseveratory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1606	0.4008
Phoneme	(Intercept)	0.1047	0.3235
Residual		0.2141	0.4627

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	1.59E+00	5.57E-01	0.501672	2.684328	6.79E+01	2.861	0.00561
RelevantDistance_using5	8.00E-03	3.26E-02	-0.0559176	0.0719136	1.16E+02	0.245	0.80667
log(WordLength*1000)	3.88E-01	8.38E-02	0.2237128	0.5522872	1.15E+02	4.629	9.72E-06

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.16274	0.4034
Phoneme	(Intercept)	0.06832	0.2614
Residual		0.51667	0.7188

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.34575	0.80442	-1.2309132	1.9224132	109.51991	0.43	0.668
RelevantDistance_using5	-0.0615	0.05019	-0.1598724	0.0368724	119.53133	-1.226	0.223
log(WordLength*1000)	-0.12402	0.1286	-0.376076	0.128036	120.17432	-0.964	0.337

Analysis 4b –infant-directed anticipatory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.12452	0.3529
Phoneme	(Intercept)	0.03142	0.1772
Residual		0.29333	0.5416

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.8427	0.67798	1.5138592	4.1715408	122.59759	4.193	5.23E-05
RelevantDistance_using5	0.02746	0.04441	-0.0595836	0.1145036	124.89544	0.618	0.537
log(WordLength*1000)	0.15381	0.11116	-0.0640636	0.3716836	123.01971	1.384	0.169

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1221	0.3494
Phoneme	(Intercept)	0.1332	0.3649
Residual		0.5259	0.7252

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	0.72535	0.91464	-1.0673444	2.5180444	110.09849	0.793	0.4295
RelevantDistance_using5	0.01871	0.05925	-0.09742	0.13484	127.55463	0.316	0.7527
log(WordLength*1000)	-0.27113	0.1487	-0.562582	0.020322	123.71646	-1.823	0.0707

Analysis 4b –infant-directed perseveratory

VOT as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.1988	0.44587
Phoneme	(Intercept)	0.006871	0.08289
Residual		0.30838	0.55532

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	2.25785	0.61141	1.0594864	3.4562136	95.24856	3.693	0.00037
RelevantDistance_using5	-0.01609	0.04266	-0.0997036	0.0675236	91.75073	-0.377	0.70695
log(WordLength*1000)	0.26463	0.09804	0.0724716	0.4567884	91.99359	2.699	0.00827

Consonant-vowel ratio as DV:

Random effects:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.22839	0.4779
Phoneme	(Intercept)	0.02599	0.1612
Residual		0.55016	0.7417

Fixed effects:

	Estimate	Std.Error	95%CI-low	95%CI-high	df	t value	Pr(>|t|)
(Intercept)	-0.31088	0.80678	-1.8921688	1.2704088	95.32249	-0.385	0.701
RelevantDistance_using5	-0.04994	0.05684	-0.1613464	0.0614664	92.77347	-0.879	0.382
log(WordLength*1000)	-0.05956	0.13047	-0.3152812	0.1961612	93.1854	-0.457	0.649

Bibliography

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press. https://www.jstor.org/stable/10.3366/j.ctvxcrw9t

Adamou, E., & Shen, X. R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, *23*(1), 53–70. https://doi.org/10.1177/1367006917709094

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

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

Aldrich, A. (2020). *Adult Early-Bilingual Speech Rhythm: Evidence from Spanish and English*. 528–532. https://doi.org/10.21437/SpeechProsody.2020-108

Amengual, M. (Ed.). (2021). *Exploring Cross-linguistic Effects and Phonetic Interactions in the Context of Bilingualism*. MDPI - Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/books978-3-0365-0967-9

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. Scopus. https://doi.org/10.1016/j.wocn.2010.09.005

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570. Scopus. https://doi.org/10.1016/j.wocn.2011.03.001

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*(4), 582–594. https://doi.org/10.1016/j.wocn.2012.05.005

Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, *21*(2), e12548. https://doi.org/10.1111/desc.12548

Antovich, D. M., & Graf Estes, K. (2020). One language or two? Navigating cross-language conflict in statistical word segmentation. *Developmental Science*, *23*(6), e12960. https://doi.org/10.1111/desc.12960

Baese-Berk, M. M., McLaughlin, D. J., & McGowan, K. B. (2020). Perception of non-native speech. *Language and Linguistics Compass*, *14*(7), e12375. https://doi.org/10.1111/lnc3.12375

Bail, A., Morini, G., & Newman, R. S. (2015). Look at the gato! Code-switching in speech to toddlers. *Journal of Child Language*, *42*(5), 1073–1101. https://doi.org/10.1017/S0305000914000695

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19*(4), 423–443. https://doi.org/10.1177/1367006913516035

Barron-Hauwaert, S. (2004). Language Strategies for Bilingual Families: The one-parent-one-language Approach. In *Language Strategies for Bilingual Families*. Multilingual Matters. https://doi.org/10.21832/9781853597169

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). Scopus. https://doi.org/10.18637/jss.v067.i01

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*(4), 847–862. https://doi.org/10.1016/j.infbeh.2013.09.001

Benders, T., Pokharel, S., & Demuth, K. (2019). Hypo-Articulation of the Four-Way Voicing Contrast in Nepali Infant-Directed Speech. *Language Learning and Development*, *15*(3), 232–254. https://doi.org/10.1080/15475441.2019.1577139

Benitez, V. L., Bulgarelli, F., Byers-Heinlein, K., Saffran, J. R., & Weiss, D. J. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. *Developmental Science*, *23*(2), e12896. https://doi.org/10.1111/desc.12896

Bergeson, T. R., Miller, R. J., & McCune, K. (2006). Mothers' Speech to Hearing-Impaired Infants and Children With Cochlear Implants. *Infancy*, *10*(3), 221–240. https://doi.org/10.1207/s15327078in1003_2

Bernstein Ratner, N. (1984). Phonological rule usage in mother-child speech. *Journal of Phonetics*, *12*(3), 245–254. https://doi.org/10.1016/S0095-4470(19)30881-2

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics*, *12*(3), 285–289. https://doi.org/10.1016/S0095-4470(19)30875-7

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language*, *11*(3), 515–522. https://doi.org/10.1017/S0305000900005924

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

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *Neuroreport*, *26*(16), 966–971. https://doi.org/10.1097/WNR.0000000000000453

Birulés, J., Pons, F., & Bosch, L. (2024). Early audiovisual language discrimination: Monolingual and bilingual infants' differences in language switch detection. *International Journal of Behavioral Development*, 01650254241252795. https://doi.org/10.1177/01650254241252795

Blackburn, A. M., & Wicha, N. Y. Y. (2022). The Effect of Code-Switching Experience on the Neural Response Elicited to a Sentential Code Switch. *Languages*, *7*(3), Article 3. https://doi.org/10.3390/languages7030178

Blanc, C., Castilla, -Earls Anny, Bunta, F., & Fulcher, -Rood Katrina. (2022). Bilingual Parents' Experiences Receiving Advice Regarding Language Use: A Qualitative Perspective. *Perspectives of the ASHA Special Interest Groups*, *7*(6), 2039–2050. https://doi.org/10.1044/2022_PERSP-22-00086

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*(5), 568–585. https://doi.org/10.1080/20445911.2013.792822

Boersma, P., & Wernick, D. (2023). *Praat: Doing phonetics by computer* (6.4.01) [Computer software]. http://www.praat.org

Bond, Z. S., & Fokes, J. (1991). Identifying foreign languages. *Proceedings of the XII International Congress of Phonetic Sciences*, 198–201.

Bond, Z. S., & Stockmal, V. (2005). *Children's abilities to distinguish novel languages*. Acoustical Society of America Conference, Vancouver.

Bond, Z. S., Stockmal, V., & Muljani, D. (1998). Learning to identify a foreign language. *Language Sciences*, *20*(4), 353–367. https://doi.org/10.1016/S0388-0001(98)00009-6

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

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

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, *64*(1), 121–130. https://doi.org/10.3758/BF03194561

Braarud, H. C., & Stormark, K. M. (2008). Prosodic Modification and Vocal Adjustments in Mothers' Speech During Face-to-face Interaction with Their Two- to Four-month-old Infants: A Double Video Study. *Social Development*, *17*(4), 1074–1084. https://doi.org/10.1111/j.1467-9507.2007.00455.x

Broersma, M., Carter, D., Donnelly, K., & Konopka, A. (2020). Triggered codeswitching: Lexical processing and conversational dynamics. *Bilingualism: Language and Cognition*, *23*(2), 295–308. https://doi.org/10.1017/S1366728919000014

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition*, *9*(1), 1–13. https://doi.org/10.1017/S1366728905002348

Buckler, H., Goy, H., & Johnson, E. K. (2018). What infant-directed speech tells us about the development of compensation for assimilation. *Journal of Phonetics*, *66*, 45–62. https://doi.org/10.1016/j.wocn.2017.09.004

Bulgarelli, F., Benitez, V., Saffran, J., Byers-Heinlein, K., & Weiss, D. J. (2017). Statistical Learning of Multiple Structures by 8-Month-Old Infants. *Proceedings of the ... Annual Boston University Conference on Language Development. Boston University Conference on Language Development*, *41*, 128–139.

Bullock, B. E., & Toribio, A. J. (2009). 8. Trying to hit a moving target: On the sociophonetics of code-switching. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary Approaches to Code Switching* (pp. 189–206). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.41.12bul

Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). *Language Dominance and Performance Outcomes in Bilingual Pronunciation*. the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference.

Bundgaard-Nielsen, R. L., O’Shannessy, C., Wang, Y., Nelson, A., Bartlett, J., & Davis, V. (2023). Two-part vowel modifications in Child Directed Speech in Warlpiri may enhance child attention to speech and scaffold noun acquisition. *Phonetica*, *80*(1–2), 1–42. https://doi.org/10.1515/phon-2022-0039

Bunta, F., Douglas, M., Dickson, H., Cantu, A., Wickesberg, J., & Gifford, R. H. (2016). Dual language versus English-only support for bilingual children with hearing loss who use cochlear implants and hearing aids. *International Journal of Language & Communication Disorders*, *51*(4), 460–472. https://doi.org/10.1111/1460-6984.12223

Burfin, S., Pascalis, O., Ruiz Tada, E., Costa, A., Savariaux, C., & Kandel, S. (2014). Bilingualism affects audiovisual phoneme identification. *Frontiers in Psychology*, *5*, 1179. https://doi.org/10.3389/fpsyg.2014.01179

Burnham, E., Gamache, J. L., Bergeson, T., & Dilley, L. (2013). *Voice-onset time in infant-directed speech over the first year and a half*. 060094–060094. https://doi.org/10.1121/1.4800072

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children’s vocabulary size. *Bilingualism: Language and Cognition*, *16*(1), 32–48. https://doi.org/10.1017/S1366728912000120

Byers-Heinlein, K. (2014). Languages As Categories: Reframing the “One Language or Two” Question in Early Bilingual Development. *Language Learning*, *64*(s2), 184–201. https://doi.org/10.1111/lang.12055

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

Byers-Heinlein, K., Jardak, A., Fourakis, E., & Lew-Williams, C. (2022). Effects of language mixing on bilingual children's word learning. *Bilingualism: Language and Cognition*, *25*(1), 55–69. https://doi.org/10.1017/S1366728921000699

Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences*, *114*(34), 9032–9037. https://doi.org/10.1073/pnas.1703220114

Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., Durrant, S., Fennell, C. T., Fiévet, A.-C., Frank, M. C., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Hamlin, J. K., Havron, N., Hernik, M., Kerr, S., Killam, H., Klassen, K., … Wermelinger, S. (2021). A Multilab Study of Bilingual Infants: Exploring the Preference for Infant-Directed Speech. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920974622. https://doi.org/10.1177/2515245920974622

Cacoullos, R. T., & Travis, C. E. (2018). *Bilingualism in the Community: Code-switching and Grammars in Contact*. Cambridge University Press. https://doi.org/10.1017/9781108235259

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or cárcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes*, *25*(5), 669–705. https://doi.org/10.1080/01690960903474912

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

Carbajal, M. J., & Peperkamp, S. (2020). Dual language input and the impact of language separation on early lexical development. *Infancy*, *25*(1), 22–45. https://doi.org/10.1111/infa.12315

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the *double phonemic boundary* in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. https://doi.org/10.1016/j.wocn.2018.07.002

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early Language Experience in a Tseltal Mayan Village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Chang, C. B. (2019). The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (1st ed., pp. 427–447). Routledge. https://doi.org/10.4324/9780429056253-16

Chasin, M. (2008). How hearing aids may be set for different languages. *Hearing Review*, *15*(11), 16–20.

Chen, L., Chao, K.-Y., & Peng, J.-F. (2007). VOT productions of word-initial stops in Mandarin and English: A cross-language study. In K.-H. Chen & B. Chen (Eds.), *ROCLING 2007 Poster Papers* (pp. 303–317). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP). https://aclanthology.org/O07-2004

Cheng, Y.-L., & Howard, D. (2008). The time cost of mixed-language processing: An investigation. *International Journal of Bilingualism*, *12*(3), 209–222. https://doi.org/10.1177/1367006908098568

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

Christophe, A., Nespor, M., Teresa Guasti, M., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, *6*(2), 211–220. https://doi.org/10.1111/1467-7687.00273

Cleave, P. L., Kay-Raining Bird, E., Trudeau, N., & Sutton, A. (2014). Syntactic bootstrapping in children with Down syndrome: The impact of bilingualism. *Journal of Communication Disorders*, *49*, 42–54. https://doi.org/10.1016/j.jcomdis.2014.02.006

Clyne, M. G. (1980). Triggering and language processing. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *34*(4), 400–406. https://doi.org/10.1037/h0081102

Comeau, L., Genesee, F., & Lapaquette, L. (2003). The Modeling Hypothesis and child bilingual codemixing. *International Journal of Bilingualism*, *7*(2), 113–126. https://doi.org/10.1177/13670069030070020101

Comeau, L., Genesee, F., & Mendelson, M. (2007). Bilingual children's repairs of breakdowns in communication. *Journal of Child Language*, *34*(1), 159–174. https://doi.org/10.1017/S0305000906007690

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, *7*(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

Cox, C., Dideriksen, C., Keren-Portnoy, T., Roepstorff, A., Christiansen, M. H., & Fusaroli, R. (2023). Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish. *Child Development*, *94*(6), 1672–1696. https://doi.org/10.1111/cdev.13950

Cristià, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America*, *128*(1), 424–434. https://doi.org/10.1121/1.3436529

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*(4), 492–504. https://doi.org/10.1016/j.wocn.2010.12.002

Dale, P. S. (1991). The Validity of a Parent Report Measure of Vocabulary and Syntax at 24 Months. *Journal of Speech, Language, and Hearing Research*, *34*(3), 565–571. https://doi.org/10.1044/jshr.3403.565

Danielson, D. K., Seidl, A., Onishi, K. H., Alamian, G., & Cristia, A. (2014). The acoustic properties of bilingual infant-directed speech. *The Journal of the Acoustical Society of America*, *135*(2), EL95–EL101. https://doi.org/10.1121/1.4862881

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

Dauer, R. M. (1987). *Phonetic and Phonological Components of Language Rhythm*. https://www.semanticscholar.org/paper/Phonetic-and-Phonological-Components-of-Language-Dauer/01b4c9c031d212683782b4de2eaee2fa98915ebb

de Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*(3), 411–424. https://doi.org/10.1017/S0142716407070221

De Houwer, A., & Bornstein, M. H. (2016). Bilingual mothers' language choice in child-directed speech: Continuity and change. *Journal of Multilingual and Multicultural Development*, *37*(7), 680–693. https://doi.org/10.1080/01434632.2015.1127929

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, *24*(3), 351–365. https://doi.org/10.1006/jpho.1996.0019

Deuchar, M., Davies, P., Herring, J. R., Couto, M. C. P., & Carter, D. (2014). 5. Building Bilingual Corpora. In *5. Building Bilingual Corpora* (pp. 93–110). Multilingual Matters. https://doi.org/10.21832/9781783091713-008

Dijkstra, T., & Heuven, W. J. B. van. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dilley, L. C., Millett, A. L., Mcauley, J. D., & Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language*, *41*(1), 155–175. https://doi.org/10.1017/S0305000912000670

Dilley, L., Lehet, M., Wieland, E. A., Arjmandi, M. K., Kondaurova, M., Wang, Y., Reed, J., Svirsky, M., Houston, D., & Bergeson, T. (2020). Individual Differences in Mothers' Spontaneous Infant-Directed Speech Predict Language Attainment in Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(7), 2453–2467. https://doi.org/10.1044/2020_JSLHR-19-00229

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*(10), 1593–1610. https://doi.org/10.1162/089892905774597245

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, *25*(2), 219–234. https://doi.org/10.1177/0142723705050286
Englund, K. T., & Behne, D. M. (2005). Infant Directed Speech in Natural Interaction—Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7
Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021). Do Infants Really Learn Phonetic Categories? *Open Mind*, *5*, 113–131. https://doi.org/10.1162/opmi_a_00046
Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104
Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679
Fernandez, C. B. (2015). *The effects of accented speech on code-switching in Spanish-English bilinguals: An electrophysiological study* [Master's thesis, Penn State University]. Penn State University LIbrary.
Fernandez, C. B., Litcofsky, K. A., & van Hell, J. G. (2019). Neural correlates of intra-sentential code-switching in the auditory modality. *Journal of Neurolinguistics*, *51*, 17–41. https://doi.org/10.1016/j.jneuroling.2018.10.004
Fish, M. S., García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics*, *63*, 19–34. https://doi.org/10.1016/j.wocn.2017.04.003
Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411. Scopus. https://doi.org/10.1121/1.400473
Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
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. https://doi.org/10.1017/9781108886901.002
Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language*, *89*, 110–137. https://doi.org/10.1016/j.jml.2015.10.001
Fritche, R., Shattuck-Hufnagel, S., & Song, J. Y. (2021). Do adults produce phonetic variants of /t/ less often in speech to children? *Journal of Phonetics*, *87*, 101056. https://doi.org/10.1016/j.wocn.2021.101056
Fuchs, R. (2016). *The acoustic correlates of stress and accent in English content and function words*. 435–439. https://doi.org/10.21437/SpeechProsody.2016-89

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

García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language*, *212*, 104890. https://doi.org/10.1016/j.bandl.2020.104890

Gasparini, L., Langus, A., Tsuji, S., & Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis. *Cognition*, *213*, 104757. https://doi.org/10.1016/j.cognition.2021.104757

Gavino, M. F., & Goldrick, M. (2023). Consequences of mixing and switching languages for retrieval and articulation. *Bilingualism: Language and Cognition*, *26*(3), 504–515. https://doi.org/10.1017/S1366728922000682

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. *Cognitive Science*, *33*(6), 1087–1116. https://doi.org/10.1111/j.1551-6709.2009.01041.x

Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, *16*(1), 161–179. https://doi.org/10.1017/S0305000900013490

Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied Psycholinguistics*, *17*(4), 427–442. https://doi.org/10.1017/S0142716400008183

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, *22*(3), 611–631. https://doi.org/10.1017/S0305000900009971

Goffman, L., & Gerken, L. (2023). A developmental account of the role of sequential dependencies in typical and atypical language learners. *Cognitive Neuropsychology*, *40*(5–6), 243–264. https://doi.org/10.1080/02643294.2023.2275837

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language Switching Makes Pronunciation Less Nativelike. *Psychological Science*, *25*(4), 1031–1036. https://doi.org/10.1177/0956797613520014

Gonzales, K., Byers-Heinlein, K., & Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they're hearing. *Cognition*, *182*, 318–330. https://doi.org/10.1016/j.cognition.2018.08.021

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*, *24*(11), 2135–2142. https://doi.org/10.1177/0956797613486485

Goodz, N. S. (1989). Parental language mixing in bilingual families. *Infant Mental Health Journal*, *10*(1), 25–44.

Gosselin, L., & Sabourin, L. (2021). Lexical-semantic processing costs are not inherent to intra-sentential code-switching: The role of switching habits. *Neuropsychologia*, *159*, 107922. https://doi.org/10.1016/j.neuropsychologia.2021.107922

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. *Laboratory Phonology 7*, 515. https://doi.org/10.1515/9783110197105.515

Graf Estes, K., & Hurley, K. (2013). Infant-Directed Prosody Helps Infants Map Sounds to Meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grammont, M. (1902). Observations sur le langage des enfants. In D. Barbelent, G. Dottin, R. Gauthiot, M. Grammont, A. Laronde, M. Niedermann, & J. Vendreyes (Eds.), *Mélanges linguistiques. Offerts à M. Antoine Meillet par ses élèves* (pp. 61–82). Klincksieck.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology (Hove, England)*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development*, *6*(6), 467–477. https://doi.org/10.1080/01434632.1985.9994221

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225–254). Lawrence Erlbaum Associates Publishers.

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

Grosjean, F. (2001). The Bilingual's Language Modes. In *One mind, two languages: Bilingual language processing* (pp. 1–22). Blackwell Publishing.

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An Example of Bilingual Flexibility. *Psychological Science*, *5*(4), 201–202. Scopus. https://doi.org/10.1111/j.1467-9280.1994.tb00501.x

Gross, M. C., Lopez, E., Buac, M., & Kaushanskaya, M. (2019). Processing of code-switched sentences by bilingual children: Cognitive and linguistic predictors. *Cognitive Development*, *52*, 100821. https://doi.org/10.1016/j.cogdev.2019.100821

Gross, M. C., Patel, H., & Kaushanskaya, M. (2021). Processing of Code-Switched Sentences in Noise by Bilingual Children. *Journal of Speech, Language, and Hearing Research : JSLHR*, *64*(4), 1283–1302. https://doi.org/10.1044/2020_JSLHR-20-00388

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When Language Switching has No Apparent Cost: Lexical Access in Sentence Context. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00278

Guzzardo Tamargo, R. E., Valdés Kroff, J. R., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language, 89*, 138–161. https://doi.org/10.1016/j.jml.2015.12.002

Hartman, K. M., Bernstein Ratner, N., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, *44*(5), 1140–1162. https://doi.org/10.1017/S0305000916000520

Henriksen, N., Coetzee, A. W., García-Amaya, L., & Fischer, M. (2021). Exploring language dominance through code-switching: Intervocalic voiced stop lenition in Afrikaans-Spanish bilinguals. *Phonetica*, *78*(3), 201–240. https://doi.org/10.1515/phon-2021-2005

Hitczenko, K., & Feldman, N. H. (2022). Naturalistic speech supports distributional learning across contexts. *Proceedings of the National Academy of Sciences*, *119*(38), e2123230119. https://doi.org/10.1073/pnas.2123230119

Jacobson, J. L., Boersma, D. C., Fields, R. B., & Olson, K. L. (1983). Paralinguistic Features of Adult Speech to Infants and Small Children. *Child Development*, *54*(2), 436–442. https://doi.org/10.2307/1129704

Jenkins, J. J., Strange, W., & Polka, L. (1995). Not everyone can tell a "rock" from a "lock": Assessing individual differences in speech perception. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 297–325). Davies-Black.

Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, *6*(3), 271–289. https://doi.org/10.1177/13670069020060030401

Johnson, E. K., & Zamuner, T. (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73–94). John Benjamins Publishing Company. https://umd.instructure.com/courses/1360351/files/76655661?module_item_id=12580370

Ju, M., & Luce, P. A. (2004). Falling on Sensitive Ears: Constraints on Bilingual Lexical Activation. *Psychological Science*, *15*(5), 314–318. https://doi.org/10.1111/j.0956-7976.2004.00675.x

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language*, *45*(5), 1035–1053. https://doi.org/10.1017/S0305000917000629

Kalashnikova, M., & Carreiras, M. (2022). Input quality and speech perception development in bilingual infants' first year of life. *Child Development*, *93*(1), e32–e46. https://doi.org/10.1111/cdev.13686

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1), 13745. https://doi.org/10.1038/s41598-018-32150-6

Kalashnikova, M., Singh, L., Tsui, A., Altuntas, E., Burnham, D., Cannistraci, R., Chin, N. B., Feng, Y., Fernández-Merino, L., Götz, A., Gustavsson, L., Hay, J., Höhle, B., Kager, R., Lai, R., Liu, L., Marklund, E., Nazzi, T., Oliveira, D. S., … Woo, P. J. (2024). The development of tone discrimination in infancy:

Evidence from a cross-linguistic, multi-lab report. *Developmental Science*, *27*(3), e13459. https://doi.org/10.1111/desc.13459
Kaushanskaya, M., Crespo, K., & Neveu, A. (2023). Does code-switching influence novel word learning? *Developmental Science*, *26*(2), e13292. https://doi.org/10.1111/desc.13292
Kay, S., & Adler, S. A. (2023). Phonemic Discrimination and Eye Movements in Infants. *Journal of Vision*, *23*(9), 5014. https://doi.org/10.1167/jov.23.9.5014
Kay-Raining Bird, E., Cleave, P., Trudeau, N., Thordardottir, E., Sutton, A., & Thorpe, A. (2005). The Language Abilities of Bilingual Children With Down Syndrome. *American Journal of Speech-Language Pathology*, *14*(3), 187–199. https://doi.org/10.1044/1058-0360(2005/019)
Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, *7*(1), 71–88. https://doi.org/10.1017/S1366728904001282
Keller, H. (2007). *Cultures of infancy* (pp. xix, 318). Lawrence Erlbaum Associates Publishers.
Kempe, V., Ota, M., & Schaeffler, S. (2024). Does child-directed speech facilitate language development in all domains? A study space analysis of the existing evidence. *Developmental Review*, *72*, 101121. https://doi.org/10.1016/j.dr.2024.101121
Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168. https://doi.org/10.1006/jpho.1996.0039
Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117–128. https://doi.org/10.1006/jpho.1997.0069
Khlystova, E. A., Chong, A. J., & Sundara, M. (2023). Phonetic variation in English infant-directed speech: A large-scale corpus analysis. *Journal of Phonetics*, *100*, 101267. https://doi.org/10.1016/j.wocn.2023.101267
Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405
Kircher, R. (2014). Thirty Years after Bill 101: A Contemporary Perspective on Attitudes Towards English and French in Montreal. *Canadian Journal of Applied Linguistics / Revue Canadienne de Linguistique Appliquee*, *17*(1), 20–50.
Kitamura, C., & Burnham, D. (2003). Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy*, *4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401_5
Koelewijn, I. M., Hoevenaars, E., & Verhagen, J. (2023). How do parents think about multilingual upbringing? Comparing OPOL parents and parents who mix languages. *Journal of Multilingual and Multicultural Development*, *0*(0), 1–17. https://doi.org/10.1080/01434632.2023.2202633
Kovács, Á. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

Kovács, Á. M., & Mehler, J. (2009b). Flexible Learning of Multiple Speech Structures in Bilingual Infants. *Science*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

Kremer-Sadlik, T. (2005). To Be or Not to Be Bilingual: Autistic children from multilingual families. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 1225–1234). Cascadilla Press.

Kremin, L. V., Alves, J., Orena, A. J., Polka, L., & Byers-Heinlein, K. (2022). Code-switching in parents' everyday speech to bilingual infants. *Journal of Child Language*, *49*(4), 714–740. https://doi.org/10.1017/S0305000921000118

Kremin, L. V., Jardak, A., Lew-Williams, C., & Byers-Heinlein, K. (2023). Bilingual children's comprehension of code-switching at an uninformative adjective. *Language Development Research*, *3*(1), 249–276.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science (New York, N.Y.)*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). York Press.

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

Kuipers, J.-R., & Thierry, G. (2010). Event-related brain potentials reveal the time-course of language change detection in early bilinguals. *NeuroImage*, *50*(4), 1633–1638. https://doi.org/10.1016/j.neuroimage.2010.01.076

Kuipers, J.-R., & Thierry, G. (2012). Event-related potential correlates of language change detection in bilingual toddlers. *Developmental Cognitive Neuroscience*, *2*(1), 97–102. https://doi.org/10.1016/j.dcn.2011.08.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). Interlingual lexical competition in a spoken sentence context: Evidence from the visual world paradigm. *Psychonomic Bulletin & Review*, *20*(5), 963–972. https://doi.org/10.3758/s13423-013-0405-4

Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, *92*, 217–233. https://doi.org/10.1016/j.jml.2016.06.010

Lavechin, M., de Seyssel, M., Métais, M., Metze, F., Mohamed, A., Bredin, H., Dupoux, E., & Cristia, A. (2024). Modeling early phonetic acquisition from child-centered audio data. *Cognition*, *245*, 105734. https://doi.org/10.1016/j.cognition.2024.105734

Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press.

Leeuw, J. R. de, 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://doi.org/10.21105/joss.05351

Lindblom, B., Hardcastle, W. J., & Marchal, A. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In *Speech Production and Speech Modeling* (Vol. 5, pp. 403–439). Springer. https://link.springer.com/chapter/10.1007/978-94-009-2037-8_16

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

Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs: Behavioral, ERP and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, *97*, 112–139. https://doi.org/10.1016/j.neuropsychologia.2017.02.002

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1–F10. https://doi.org/10.1111/1467-7687.00275

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

López Velarde, M., & Simonet, M. (2020). The Perception of Postalveolar English Obstruents by Spanish Speakers Learning English as a Foreign Language in Mexico. *Languages*, *5*(2), 27. https://doi.org/10.3390/languages5020027

Lorch, M. P., & Meara, P. (1989). How people listen to languages they don't know. *Language Sciences*, *11*(4), 343–353. https://doi.org/10.1016/0388-0001(89)90025-9

Lorch, M. P., & Meara, P. (1995). Can people discriminate languages they don't know? *Language Sciences*, *17*(1), 65–71. https://doi.org/10.1016/0388-0001(95)00003-E

Lovcevic, I., Kalashnikova, M., & Burnham, D. (2020). Acoustic features of infant-directed speech to infants with hearing loss. *The Journal of the Acoustical Society of America*, *148*(6), 3399–3416. https://doi.org/10.1121/10.0002641

Ludusan, B., Mazuka, R., & Dupoux, E. (2021). Does Infant-Directed Speech Help Phonetic Learning? A Machine Learning Investigation. *Cognitive Science*, *45*(5), e12946. https://doi.org/10.1111/cogs.12946

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. https://doi.org/10.1016/j.bandc.2008.08.019

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development : The Official Journal of the Society for Language Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

MacDonald, M. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00226

Macleod, A. a. N., & Stoel-Gammon, C. (2009). The use of voice onset time by early bilinguals to distinguish homorganic stops in Canadian English and Canadian French. *Applied Psycholinguistics*, *30*(1), 53–77. https://doi.org/10.1017/S0142716408090036

MacLeod, A. A., & Stoel-Gammon, C. (2010). Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders*. https://doi.org/10.1080/14769670500066313

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Volume 2 Perception*. Academic Press.

Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, *20*(1), e12425. https://doi.org/10.1111/desc.12425

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(2), 173–193. https://doi.org/10.1017/S0142716403000092

Marklund, E., Marklund, U., & Gustavsson, L. (2021). An Association Between Phonetic Complexity of Infant Vocalizations and Parent Vowel Hyperarticulation. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.693866

Marks, E. A., Bond, Z. S., & Stockmal, V. (1999). *THE EFFECT OF PROFICIENCY IN A SPECIFIC FOREIGN LANGUAGE ON THE ABILITY TO IDENTIFY A NOVEL FOREIGN LANGUAGE*. 133–135.

Marks, E. A., Bond, Z. S., & Stockmal, V. (2001). *Language Experience and the Representation of Phonology in an Unknown Language*. https://eric.ed.gov/?id=ED458823

Marks, E. A., Bretón, N. O., Bond, Z. S., Stockmal, V., & McKenna, L. (2009). Monolingual & bilingual children's abilities to distinguish novel languages. In J. Valenzuela (Ed.), *Trends in Cognitive Linguistics: Theoretical and Applied Models* (pp. 279–292). Peter Lang.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Martin, C. D., & Nozari, N. (2021). Language control in bilingual production: Insights from error rate and error type in sentence production. *Bilingualism: Language and Cognition*, *24*(2), 374–388. https://doi.org/10.1017/S1366728920000590

Matsumoto, N., Suzuki, N., Iwasaki, S., Ishikawa, K., Tsukiji, H., Higashino, Y., Tabuki, T., & Nakagawa, T. (2018). Language-specific strategy for programming hearing aids—A double-blind randomized controlled crossover study. *Auris Nasus Larynx*, *45*(4), 686–692. https://doi.org/10.1016/j.anl.2017.11.007

Matusevych, Y., Schatz, T., Kamper, H., Feldman, N. H., & Goldwater, S. (2023). Infant Phonetic Learning as Perceptual Space Learning: A Crosslinguistic Evaluation of Computational Models. *Cognitive Science*, *47*(7). Scopus. https://doi.org/10.1111/cogs.13314

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi*. 498–502. https://doi.org/10.21437/Interspeech.2017-1386

McDaniel, J., Benítez-Barrera, C. R., Soares, A. C., Vargas, A., & Camarata, S. (2019). Bilingual Versus Monolingual Vocabulary Instruction for Bilingual Children with Hearing Loss. *The Journal of Deaf Studies and Deaf Education*, *24*(2), 142–160. https://doi.org/10.1093/deafed/eny042

McMurray, B. (2023). The acquisition of speech categories: Beyond perceptual narrowing, beyond unsupervised learning and beyond infancy. *Language, Cognition and Neuroscience*, *38*(4), 419–445. https://doi.org/10.1080/23273798.2022.2105367

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203–229. https://doi.org/10.1207/s15327078in0602_4

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, *129*(2), 362–378. https://doi.org/10.1016/j.cognition.2013.07.015

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

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of Cognate Status on Word Comprehension in Second Language Learners: An ERP Investigation. *Journal of Cognitive Neuroscience*, *23*(7), 1634–1647. https://doi.org/10.1162/jocn.2010.21463

Migration Policy Institute. (2021, May 18). *Young Dual Language Learners in the United States and by State*. Migrationpolicy.Org. https://www.migrationpolicy.org/programs/data-hub/charts/us-state-profiles-young-dlls

Miller, J. L., Green, K. P., & Reeves, A. (2009). 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

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, *166*, 84–93. https://doi.org/10.1016/j.cognition.2017.05.003

Mok, P. P. K. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, *2*, 148–154.

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm Class Native Language Discrimination Abilities of Basque-Spanish Monolingual and Bilingual Infants at 3.5 Months of Age. *Infancy*, *19*(3), 326–337. https://doi.org/10.1111/infa.12041

Molnar, M., Ibáñez-Molina, A., & Carreiras, M. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language*, *81*, 91–104. https://doi.org/10.1016/j.jml.2015.01.002

Montanari, S., Steffman, J., & Mayr, R. (2023). Stop voicing perception in the societal and heritage language of Spanish-English bilingual preschoolers: The role of age, input quantity and input diversity. *Journal of Phonetics*, *101*. Scopus. https://doi.org/10.1016/j.wocn.2023.101276

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

Morini, G., & Newman, R. S. (2019). Dónde está la ball? Examining the effect of code switching on bilingual children's word recognition. *Journal of Child Language*, *46*(6), 1238–1248. https://doi.org/10.1017/S0305000919000400

Muldner, K., Hoiting, L., Sanger, L., Blumenfeld, L., & Toivonen, I. (2019). The phonetics of code-switched vowels. *International Journal of Bilingualism*, *23*(1), 37–52. https://doi.org/10.1177/1367006917709093

Muru, A. M., & Lee, S. A. S. (2017). Development of Phonetic Categories of Stop Consonants in Spanish-English Bilingual Children. *Clinical Archives of Communication Disorders*, *2*(1), 60–68. https://doi.org/10.21849/cacd.2017.00066

Muthusamy, Y. K., Jain, N., & Cole, R. A. (1994). Perceptual benchmarks for automatic language identification. *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, *i*, I/333-I/336 vol.1. https://doi.org/10.1109/ICASSP.1994.389288

Muylle, M., Assche, E. V., & Hartsuiker, R. J. (2022). Comparing the cognate effect in spoken and written second language word production. *Bilingualism: Language and Cognition*, *25*(1), 93–107. https://doi.org/10.1017/S1366728921000444

Muysken, P. (2000). *Bilingual Speech: A Typology of Code-Mixing*. Cambridge University Press.

mybeautifulblunder [@mybeautifulblunder] (Director). (2023, August 14). *Raising bilingual kids with my English only speaking partner #codeswitching #raisingbilingualchildren #raisingbilingualkids #raisingbilingualbabies #bilingualbaby #bilingualparenting [Video]* [TikTok]. https://www.tiktok.com/t/ZPRwB1HVf/

Myers-Scotton, C. (1997). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford University Press.

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

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, *139*(3), 1272–1281. https://doi.org/10.1121/1.4944634

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

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language*, *43*(1), 1–19. https://doi.org/10.1006/jmla.2000.2698

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047. https://doi.org/10.1016/j.dr.2022.101047

Netelenbos, N., & Li, F. (2013). *The production and perception of voice onset time in English-speaking children enrolled in a French immersion program*. 2380–2384. https://doi.org/10.21437/Interspeech.2013-555

Ng, S., & Wicha, N. Y. Y. (2013). Meaning first: A case for language-independent access to word meaning in the bilingual brain. *Neuropsychologia*, *51*(5), 850–863. https://doi.org/10.1016/j.neuropsychologia.2013.01.017

Nilep, C. (2006). “Code Switching” in Sociocultural Linguistics. *Colorado Research in Linguistics*. https://doi.org/10.25810/hnq4-jv62

Økland, K. E. (2021). *Infant-Directed Speech in Norwegian Fathers and Mothers: Relations to Gender and Gender-Role Attitudes* [Master thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2776119

Olson, D. J. (2012). The phonetics of insertional code-switching: Suprasegmental analysis and a case for hyper-articulation. *Linguistic Approaches to Bilingualism*, *2*(4), 439–457. Scopus. https://doi.org/10.1075/lab.2.4.05ols

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*(6), 407–420. Scopus. https://doi.org/10.1016/j.wocn.2013.07.005

Olson, D. J. (2016). The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, *46*(3), 263–285. Scopus. https://doi.org/10.1017/S0025100315000468

Olson, D. J. (2019). Phonological processes across word and language boundaries: Evidence from code-switching. *Journal of Phonetics*, *77*. Scopus. https://doi.org/10.1016/j.wocn.2019.100937

Olson, D. J. (2020). Short-Term Sources of Cross-Linguistic Phonetic Influence: Examining the Role of Linguistic Environment. *Languages*, *5*(4), Article 4. https://doi.org/10.3390/languages5040043

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019). Reliability of the Language Environment Analysis Recording System in Analyzing French-English Bilingual Speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(7), 2491–2500. https://doi.org/10.1044/2019_JSLHR-L-18-0342

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

Peynircioğlu, Z. F., & Durgunoğlu, A. Y. (2002). 15 Code-switching in preschool bilingual children. In R. R. Heredia & J. Altarriba (Eds.), *Advances in Psychology* (Vol. 134, pp. 339–356). North-Holland. https://doi.org/10.1016/S0166-4115(02)80018-8

Phillips, J. R. (1973). Syntax and Vocabulary of Mothers' Speech to Young Children: Age and Sex Comparisons. *Child Development*, *44*(1), 182–185. https://doi.org/10.2307/1127699

Piazza, G., Martin, C. D., & Kalashnikova, M. (2022). The Acoustic Features and Didactic Function of Foreigner-Directed Speech: A Scoping Review. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2896–2918. https://doi.org/10.1044/2022_JSLHR-21-00609

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, *52*, 121–137. https://doi.org/10.1016/j.wocn.2015.07.004

Piccinini, P., & Garellek, M. (2014). Prosodic Cues to Monolingual versus Code-switching Sentences in English and Spanish. *Speech Prosody 2014*, 885–889. https://doi.org/10.21437/SpeechProsody.2014-166

Piller, I., & Gerber, L. (2021). Family language policy between the bilingual advantage and the monolingual mindset. *International Journal of Bilingual Education and Bilingualism*, *24*(5), 622–635. https://doi.org/10.1080/13670050.2018.1503227

Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, *19*(5), 1023–1041. https://doi.org/10.1017/S1366728915000322

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, *39*(3), 362–374. https://doi.org/10.1016/j.wocn.2011.03.004

Poplack, S. (1980). *Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching1*. *18*(7–8), 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., & Lew-Williams, C. (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, *22*(4), e12794. https://doi.org/10.1111/desc.12794

Potter, C. E., & Saffran, J. R. (2015). The role of experience in children's discrimination of unfamiliar languages. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01587
Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to Represent a Multi-Context Environment: More than Detecting Changes. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00228
Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, *2*, 85–115.
Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, *3*.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X
Robertson, S., von, H. D., & Hay, J. S. (2013). The Effect of Hearing Loss on the Perception of Infant- and Adult-Directed Speech. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1108–1119. https://doi.org/10.1044/1092-4388(2012/12-0110)
Rosslund, A., Mayor, J., Óturai, G., & Kartushina, N. (2022). Parents' hyper-pitch and low vowel category variability in infant-directed speech are associated with 18-month-old toddlers' expressive vocabulary. *Language Development Research*, *2*(1), 223–267. https://doi.org/10.34842/2022.0547
Rosslund, A., Varjola, N. E. H., Mayor, J., & Kartushina, N. (2023). *Longitudinal changes in consonant production in infant-directed speech and infants' early speech production from 6 to 12 months*. OSF. https://doi.org/10.31234/osf.io/8y9wx
Ruan, Y., Byers-Heinlein, K., Orena, A. J., & Polka, L. (2023). Mixed-language input and infant volubility: Friend or foe? *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/S1366728923000287
Salig, L. K., Valdés Kroff, J. R., Slevc, L. R., & Novick, J. M. (2023). Linking frequency to bilingual switch costs during real-time sentence comprehension. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/S1366728923000366
Sánchez, M. J., & Pérez-García, E. (2020). Code-Switching, Language Emotionality and Identity in Junot Díaz's "Invierno." *Atlantis*, *42*(2), 99–118.
Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051
Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, *118*(7), e2001844118. https://doi.org/10.1073/pnas.2001844118
Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of

distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434
Schott, E., Mastroberardino, M., Fourakis, E., Lew-Williams, C., & Byers-Heinlein, K. (2021). Fine-tuning language discrimination: Bilingual and monolingual infants' detection of language switching. *Infancy*, *26*(6), 1037–1056. https://doi.org/10.1111/infa.12429
Schott, E., Tamayo, M. P., & Byers-Heinlein, K. (2023). Keeping track of language: Can monolingual and bilingual infants associate a speaker with the language they speak? *Infant and Child Development*, *32*(3), e2403. https://doi.org/10.1002/icd.2403
Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004
Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. *Open Mind*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002
Segal, O., & Kishon-Rabin, L. (2011). Listening preference for child-directed speech versus nonspeech stimuli in normal-hearing and hearing-impaired infants after cochlear implantation. *Ear and Hearing*, *32*(3), 358–372. https://doi.org/10.1097/AUD.0b013e3182008afc
Shafer, V. L., Shucard, D. W., & Jaeger, J. J. (1999). Electrophysiological indices of cerebral specialization and the role of prosody in language acquisition in 3-month-old infants. *Developmental Neuropsychology*, *15*(1), 73–109. https://doi.org/10.1080/87565649909540740
Shen, A., Gahl, S., & Johnson, K. (2020). Didn't hear that coming: Effects of withholding phonetic cues to code-switching. *Bilingualism: Language and Cognition*, *23*(5), 1020–1031. https://doi.org/10.1017/S1366728919000877
Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism (Cambridge, England)*, *16*(2). https://doi.org/10.1017/S1366728912000466
Simonet, M. (2016). The Phonetics and Phonology of Bilingualism. In Oxford Handbooks Editorial Board (Ed.), *Oxford Handbook Topics in Linguistics* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199935345.013.72
Singh, L., Loh, D., & Xiao, N. G. (2017). Bilingual Infants Demonstrate Perceptual Flexibility in Phoneme Discrimination but Perceptual Constraint in Face Discrimination. *Frontiers in Psychology*, *8*, 1563. https://doi.org/10.3389/fpsyg.2017.01563
Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719
Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*(4), 869–902. https://doi.org/10.1017/S0305000908008763

Spinelli, M., & Mesman, J. (2018). The Regulation of Infant Negative Emotions: The Role of Maternal Sensitivity and Infant-Directed Speech Prosody. *Infancy*, *23*(4), 502–518. https://doi.org/10.1111/infa.12237

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

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Starreveld, P. A., Groot, A. M. B. D., Rossmark, B. M. M., & Hell, J. G. V. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, *17*(2), 258–276. https://doi.org/10.1017/S1366728913000308

Stockmal, V. (1995). *Discrimination of unknown foreign languages in spoken utterances: A developmental study* [Unpublished master's thesis]. Ohio University.

Stockmal, V., Moates, D. R., & Bond, Z. S. (1998). Same talker, different language. *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 0164-0. https://doi.org/10.21437/ICSLP.1998-204

Stockmal, V., Muljani, D., & Bond, Z. (1994). Can children identify samples of foreign languages as same or different? *Language Sciences*, *16*(2), 237–252. https://doi.org/10.1016/0388-0001(94)90002-7

Stockmal, V., Muljani, D., & Bond, Z. S. (1996). *Perceptual features of unknown foreign languages as revealed by multi-dimensional scaling*. 1748–1751. https://doi.org/10.21437/ICSLP.1996-444

Stoehr, A., Benders, T., Hell, J. G. V., & Fikkert, P. (2018). Heritage language exposure impacts voice onset time of Dutch–German simultaneous bilingual preschoolers. *Bilingualism: Language and Cognition*, *21*(3), 598–617. https://doi.org/10.1017/S1366728917000116

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*(1), 97–114. https://doi.org/10.1017/S1366728905002403

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

Sundara, M., Ward, N., Conboy, B., & Kuhl, P. K. (2020). Exposure to a second language in infancy alters speech production. *Bilingualism: Language and Cognition*, *23*(5), 978–991. https://doi.org/10.1017/S1366728919000853

Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one- year-olds. *Lund University, Dept. of Linguistics Working Papers*, *49*, 148–151.

Sundberg, U., & Lacerda, F. (1999). Voice Onset Time in Speech to Infants and Adults. *Phonetica*, *56*(3–4), 186–199. https://doi.org/10.1159/000028450

Synnestvedt, A., Bernstein Ratner, N., & Newman, R. (2010). Voice onset time in infant-directed speech at 7.5 and 11 months. *The Journal of the Acoustical Society of America*, *127*, 1853. https://doi.org/10.1121/1.3384386

The ManyBabies Consortium, Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., Bergelson, E., Bergmann, C., Black, A. K., Blything, R., Böhland, M. P., Bolitho, P., Borovsky, A., Brady, S. M., Braun, B., Brown, A., Byers-Heinlein, K., Campbell, L. E., … Soderstrom, M. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Tomić, A., & Valdés Kroff, J. R. (2022). Expecting the unexpected: Code-switching as a facilitatory cue in online sentence processing. *Bilingualism: Language and Cognition*, *25*(1), 81–92. https://doi.org/10.1017/S1366728921000237

Travis, C. E., & Cacoullos, R. T. (2013). Making Voices Count: Corpus Compilation in Bilingual Communities. *Australian Journal of Linguistics*, *33*(2), 170–194. https://doi.org/10.1080/07268602.2013.814529

Trelles, M. P., & Castro, K. (2019). Bilingualism in Autism Spectrum Disorder: Finding Meaning in Translation. *Journal of the American Academy of Child and Adolescent Psychiatry*, *58*(11), 1035. https://doi.org/10.1016/j.jaac.2019.05.027

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*(4), 2285–2294. https://doi.org/10.1121/1.2338290

Tsui, A. S. M., Erickson, L. C., Mallikarjunn, A., Thiessen, E. D., & Fennell, C. T. (2021). Dual language statistical word segmentation in infancy: Simulating a language-mixing bilingual environment. *Developmental Science*, *24*(3), e13050. https://doi.org/10.1111/desc.13050

Tsui, R. K.-Y., Kosie, J. E., Fibla, L., Lew-Williams, C., & Byers-Heinlein, K. (2023). Patterns of language switching and bilingual children's word learning: An experiment across two communities. *Translational Issues in Psychological Science*, *9*(4), 323–337. https://doi.org/10.1037/tps0000353

Tsui, R. K.-Y., Tong, X., & Chan, C. S. K. (2019). Impact of language dominance on phonetic transfer in Cantonese-English bilingual language switching. *Applied Psycholinguistics*, *40*(1), 29–58. Scopus. https://doi.org/10.1017/S0142716418000449

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7. https://doi.org/10.1016/j.specom.2006.10.003

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does Bilingualism Change Native-Language Reading? Cognate Effects in a Sentence Context. *Psychological Science*, *20*(8), 923–927.

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789. https://doi.org/10.3758/BF03196335

Van Hell, J. G., Fernandez, C. B., Kootstra, G. J., Litcofsky, K. A., & Ting, C. Y. (2018). Electrophysiological and experimental-behavioral approaches to the study of intra-sentential code-switching. *Linguistic Approaches to Bilingualism*, *8*(1), 134–161. https://doi.org/10.1075/lab.16010.van

Van Hell, J. G., Litcofsky, K. A., & Fernandez, C. B. (2015). Intra-sentential code-switching: Cognitive and neural approaches. In J. W. Schweiter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 459–482). Cambridge University Press. https://repository.ubn.ru.nl/handle/2066/150700

Van Hell, J. G., & Witteman, M. J. (2009). The neurocognition of switching between languages: A review of electrophysiological studies. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to code switching* (pp. 53–84). John Benjamins Publishing Company. https://www.mpi.nl/publications/item68482/neurocognition-switching-between-languages-review-electrophysiological

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01000

Verhagen, J., Kuiken, F., & Andringa, S. (2022). Family language patterns in bilingual families and relationships with children's language outcomes. *Applied Psycholinguistics*, *43*(5), 1109–1139. https://doi.org/10.1017/S0142716422000297

Wang, C., Xu, Y., & Zhang, J. (2023). Functional timing or rhythmical timing, or both? A corpus study of English and Mandarin duration. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.869049

Wang, M., Jegathesan, T., Young, E., Huber, J., & Minhas, R. (2018). Raising Children with Autism Spectrum Disorders in Monolingual vs Bilingual Homes: A Scoping Review. *Journal of Developmental & Behavioral Pediatrics*, *39*(5), 434. https://doi.org/10.1097/DBP.0000000000000574

Wang, Y., Bergeson, T. R., & Houston, D. M. (2017). Infant-Directed Speech Enhances Attention to Speech in Deaf Infants With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3321–3333. https://doi.org/10.1044/2017_JSLHR-H-17-0149

Wang, Y., Bergeson, T. R., & Houston, D. M. (2018). Preference for Infant-Directed Speech in Infants With Hearing Aids: Effects of Early Auditory Experience. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2431–2439. https://doi.org/10.1044/2018_JSLHR-H-18-0086

Wang, Y., Houston, D. M., & Seidl, A. (2018). Acoustic properties of infant-directed speech. In S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 93–115). Oxford University Press.

Ward, R., & Sanoudaki, E. (2021). Language profiles of Welsh-English bilingual children with Down syndrome. *Journal of Communication Disorders*, *93*, 106126. https://doi.org/10.1016/j.jcomdis.2021.106126

Weber, A., Betta, A. M. D., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34–51. https://doi.org/10.1016/j.wocn.2014.05.002

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

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

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

White, L., Floccia, C., Goslin, J., & Butler, J. (2014). Utterance-Final Lengthening Is Predictive of Infants' Discrimination of English Accents. *Language Learning*, *64*(s2), 27–44. https://doi.org/10.1111/lang.12060

White, L., Luche, C. D., & Floccia, C. (2016). Five-month-old infants' discrimination of unfamiliar languages does not accord with "rhythm class." *Speech Prosody 2016*, 567–571. https://doi.org/10.21437/SpeechProsody.2016-116

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

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wieland, E. A., Burnham, E. B., Kondaurova, M., Bergeson, T. R., & Dilley, L. C. (2015). Vowel space characteristics of speech directed to children with and without hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(2), 254–267. https://doi.org/10.1044/2015_JSLHR-S-13-0250

Williams, A., Srinivasan, M., Liu, C., Lee, P., & Zhou, Q. (2020). Why do bilingual code-switch when emotional? *Emotion (Washington, D.C.)*, *20*(5), 830–841. https://doi.org/10.1037/emo0000568

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *The Journal of Neuroscience*, *30*(22), 7646–7651. https://doi.org/10.1523/JNEUROSCI.1602-10.2010

Xu, Y. (1997). Contextual Tonal Variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Yacovone, A., Moya, E., & Snedeker, J. (2021). Unexpected words or unexpected languages? Two ERP effects of code-switching in naturalistic discourse. *Cognition*, *215*, 104814. https://doi.org/10.1016/j.cognition.2021.104814

Yamamoto, M. (2001). *Language Use in Interlingual Families: A Japanese-English Sociolinguistic Study*. Multilingual Matters. https://doi.org/10.21832/9781853595417

Zacharaki, K., & Sebastian-Galles, N. (2021). The ontogeny of early language discrimination: Beyond rhythm. *Cognition*, *213*, 104628. https://doi.org/10.1016/j.cognition.2021.104628