Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Tamar Gilad April 10, 2024

Is There a Bilingual Advantage in Talker Identification?

by

Tamar Gilad

Dr. Lynne Nygaard
Adviser

Neuroscience and Behavioral Biology

Dr. Lynne Nygaard

Adviser

Dr. Andrea Roeser

Committee Member

Dr. Brianna L. Yamasaki

Committee Member

2024

Is There a Bilingual Advantage in Talker Identification?

By

Tamar Gilad

Dr. Lynne Nygaard

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Neuroscience and Behavioral Biology

2024

Abstract

Is There a Bilingual Advantage in Talker Identification?
By Tamar Gilad

Listeners are better at identifying talkers that speak in a familiar language rather than an unfamiliar or foreign one. Although this language familiarity effect suggests that listeners' ability to recognize voices depends on exposure to language-specific regularities, the mechanisms that underlie the ability to identify voices across languages are unclear. Here, we examine whether a bilingual advantage exists for identifying talkers who speak in a familiar or unfamiliar foreign accent. In an online experiment, monolingual English speakers and bilingual (Spanish/English and other language/English) participants were exposed to talkers producing Spanish-accented English, using an implicit task focusing on the verbal content of their speech. At test, participants were then asked to learn to identify Spanish-accented talkers' voices, which were either the same or different from the voices heard during the exposure phase. Results indicate that all participants were able to learn to identify the talkers' voices at test. However, monolinguals outperformed both groups of bilinguals in identifying the Spanish-accented talkers. Spanish-English bilinguals performed better at talker identification tasks when tested on familiar voices, exhibiting an advantage for previous exposure to the voices, while monolinguals and bilinguals of other languages did not significantly benefit from previous exposure to the voices. These findings suggest that talker identification may depend on familiarity with specific aspects of language structure, such as vocabulary or phonological form. However, a general bilingual advantage for talker identification may not exist.

Is There a Bilingual Advantage in Talker Identification?

By

Tamar Gilad

Dr. Lynne Nygaard

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science Honors

Neuroscience and Behavioral Biology

2024

Acknowledgements

I would like to express my sincerest gratitude to Dr. Lynne Nygaard for her indispensable mentorship, unwavering support, and critical feedback throughout the year. I will be forever grateful to have had the opportunity to learn under such an incredible and knowledgeable role model. Special thanks to the members of the Speech and Language Perception Laboratory, especially Dr. Kate Revill and graduate students, Ana Maria Hoffmann, Leonardo Michelini, and Siyu Lin, for their valuable advice and assistance with technology. To my committee members, Dr. Brianna L. Yamasaki and Dr. Andrea Roeser, thank you for your incredible feedback and support. Lastly, I would like to extend my gratitude for Emory University whose resources and facilities were essential for my thesis completion.

Contents

IS THERE A BILINGUAL ADVANTAGE IN TALKER IDENTIFICATION?

Introduction

As social creatures, human listeners are exposed to a variety of different voices on a daily basis. Each voice differs from another as a result of individual structural differences of the vocal tract (shape and size) and idiosyncratic styles of production or speech habits (Fant, 1973; Mullennix & Pisoni, 1990; Peterson & Barney, 1952). Talkers might speak in the same or different language, with a foreign accent or an unfamiliar dialect, as well as in a variety of conversational contexts. In general, despite this variability, individuals readily learn to identify talkers' voices through exposure, and can recognize a familiar voice as opposed to an unfamiliar voice, a phenomenon known as the *voice familiarity effect* (Holmes et al, 2021). Although the length of exposure needed to learn to identify a voice may vary, listeners can become familiar with a voice within a short time. Holmes et al., (2021) demonstrated that listeners were able to recognize a voice with just 10 minutes of exposure. Furthermore, voices which participants are trained to recognize over the course of several days, are perceived as more intelligible by participants than novel voices (Kreitewolf et al., 2017; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Yonan & Sommers, 2000). Yet, given the multitude of ways in which voices can vary and the conditions in which some voices are easier to learn and perceive than others, it can also be challenging for a listener to identify a talker's voice. As learning to identify a talker's voice is an essential task for interpersonal communication and learning, it is important to understand what factors aid a listener in identifying specific voices over others.

Talker Identification: Language Familiarity Effects

Previous research suggests that human listeners are better able to identify voices speaking in a familiar language, accent, or dialect (Bregman & Creel, 2014; Perrachione 2019; Perrachione et al., 2011; Perrachione et al., 2009; Stevenage et al., 2012; Thompson, 1987; Winters et al., 2008). For example, when exposed to both English and German talkers, native English monolingual speakers were better at identifying talkers that spoke in English as compared to talkers that spoke in German (Perrachione et al., 2009). This finding exemplifies the *language-familiarity effect* (Goggin et al., 1991): a phenomenon in which listeners are better at distinguishing and identifying talkers that speak in a familiar language rather than a foreign one. The role that language familiarity plays in speech perception and talker identification is strongly supported by studies which compare performance in talker identification tasks between bilingual and monolingual individuals (Bregman & Creel, 2014; Orena et al., 2019). When tasked with identifying talkers who spoke both Korean and English, Bregman and Creel (2014) found that Korean-English bilinguals exhibited higher learning rates for recognizing talkers that spoke in Korean (their first language), than monolinguals who were unfamiliar with Korean. This benefit in identifying voices speaking in a familiar language has been further supported by brain imaging studies, which suggest that there is enhanced activation in the left hemisphere neural systems when individuals are listening to talkers speaking in a familiar language as opposed to a foreign one (Perrachione & Wong 2007; Perrachione et al., 2009).

Several studies also report that listeners perform better in voice identification tasks when talkers speak with the same accent as the listeners', even when speaking in their non-native language, suggesting that the language-familiarity effect also applies to familiarity with accents (Goggin et al., 1991; Stevenage et al., 2012). Speech spoken by non-native speakers is filtered through the phonological structure of the speakers' native language (Flege et al., 1999; Flege et al., 1997; Flege & Fletcher, 1992), creating a consistent source of variation associated with that particular accent. Presumably, speakers that share a native language and phonological structure would be able to use that familiarity when identifying talkers with the same accent.

Based on the literature, it appears as though listeners use a combination of indexical (non-linguistic) and linguistic cues in speech in order to both identify talkers and to understand speech (Nygaard et al., 1994; Perrachione et al., 2011). This finding has been supported through experiments such as those performed by Winters et al. (2008), in which listeners must identify and discriminate between bilingual talkers speaking two different languages. Winters et al. (2008) concluded that while individuals use mainly language-independent cues when identifying talkers speaking in a foreign language, participants are better able to perceive and discriminate talkers who are speaking in a familiar language, suggesting the use of linguistic cues to talker identity. Thus, individuals often use a mixture of both language independent and dependent cues when identifying a voice. However, there is little to no consensus on the underlying mechanisms responsible for these effects.

In an effort to better understand why talker identification would depend on language dependent cues, Perrachione (2019) reviewed over 30 years worth of research pertaining to the language familiarity effect and concluded that it "appears to depend primarily on familiarity with the phonological system of language and memory for words," (p.14-15). In other words, listeners who are familiar with the way speech sounds are produced in a specific phonological form, are better able to perceive an individual talker's unique speech productions and more easily identify a talker. One study of Perrachione and colleagues (2011) supports this claim. Individuals who were dyslexic were compared with individual who were non-dyslexic on a voice recognition task. Talkers either spoke in languages that were familiar or unfamiliar to all participants (Perrachione et al., 2011). Given that one of the defining characteristics of dyslexia is impaired phonological processing, individuals who are dyslexic were hypothesized to have difficulty benefiting from talkers speaking in a familiar language because familiarity with phonology aids in voice recognition and identification (Bradley & Bryant, 1983; Gabrieli, 2009). As predicted, participants with dyslexia performed equally well as participants without dyslexia when recognizing talkers speaking in an unfamiliar language, but exhibited a disadvantage compared to non-dyslexic individuals for recognizing voices speaking in a familiar language (Perrachione et al., 2011). The dyslexics' impaired performance suggests that familiarity with phonology does indeed assist individuals in recognizing talkers as the dyslexic individuals were presumably unable to access their stored phonological knowledge for the familiar language. However, the mechanisms which may aid in identifying talkers speaking in a familiar language may not be limited to familiarity with phonology. There is need for more

research to explore if more than one mechanism is at play and whether they are implemented differently when identifying talkers speaking in familiar versus unfamiliar language. Given their broader language exposure and familiarity with the sound structure of more than one language, bilinguals serve as an optimal subject group for better understanding what mechanisms are at play, the possible interactions between mechanisms, and the conditions under which different mechanisms are used.

Bilinguals: Better Understanding Talker Identification Mechanisms

Several studies suggest that infants raised as bilingual display an advantage over monolinguals when it comes to processing auditory and visual stimuli as well as performing certain cognitive tasks (Brito & Barr, 2014; Liu & Kager, 2017; Sebastián-Gallés et al., 2012). Given evidence for these cognitive advantages for bilinguals, as well as their enhanced exposure to multiple languages, and experience switching between languages, bilinguals may outperform monolinguals in talker identification tasks. As reviewed above, several studies indicate that information about the structure of language allows a listener to better recognize a specific voice (Perrachione, 2019). Bilinguals' experience with the linguistic structure of more than one spoken language may facilitate the processing of talker identity. A helpful starting point to evaluate what mechanisms might be responsible for any talker identification advantage is research examining how listeners perform in talker identification tasks when they hear the same talker speak multiple languages, and how this performance varies given their own language experience.

Possible Bilingual Advantage in Talker Identification Tasks

Given that individuals are better at identifying voices in a language or accent that is familiar to them, one would expect that bilinguals would outperform monolinguals when tasked with identifying voices speaking in a language that is familiar to them but unfamiliar to monolinguals. Orena et al. (2019) investigated whether bilinguals outperform monolinguals in talker identification tasks in which the talkers speak two languages. More specifically, English monolinguals and French-English bilinguals were tasked with identifying talkers who spoke English and then switched to French. As supported by the language familiarity effect, bilinguals were better at identifying voices following a language switch (regardless if it was English to French or French to English), than monolinguals. Based on these findings, Orena et al. (2019) propose two possible mechanisms for bilinguals' enhanced performance. One was that bilinguals may have better cognitive control (higher sensitivity to systematic differences when hearing voices). On this view, bilinguals are better able to focus on the talker-specific characteristics of speech more generally. Second was that bilinguals may have higher receptivity to altered phonological structure due to their language experience, and exposure to speakers of different languages and accents. Familiarity with phonetic structure as a possible mechanism used by bilinguals during talker identification tasks is supported by other studies which suggest a positive relationship between discriminating second language phonemes and recognizing voices among adult bilinguals (Díaz et al., 2022). This advantage has been identified in both behavioral and brain electrophysiological tests (Díaz et al., 2022). Familiarity with phonetic structure may indeed have aided French-English bilinguals during a talker identification task when

presented with French and English auditory speech stimuli. However, because Orena et al. (2019) only recruited a specific group of bilinguals and did not compare them with other types of bilinguals that had less familiarity with French phonetic structure (such as English-Mandarin bilinguals), it is difficult to conclude that it is familiarity with a specific phonological system that facilitated better performance for bilinguals or rather a more general sensitivity to the sound structure of language or cognitive processing. In order to argue whether these mechanisms are particular to certain bilingual groups or not, studies must compare performance among multiple bilingual subject groups, which is what the present study aimed to do.

Although Orena et al.'s (2019) study demonstrates that bilinguals are better at identifying voices that speak in two different languages, both of which are familiar to them, bilinguals also appear to display an advantage over monolinguals in identifying voices that speak in a language that is unfamiliar to them. Fecher and Johnson (2018a,b) found differences in processing putatively non-linguistic aspects of speech, such as may be used in talker identification, among infants being raised in bilingual and monolingual environments during infancy. In one study, monolingual and bilingual 9-month-olds were tested on identification of talkers speaking in Spanish, a language unfamiliar to both participant groups (Fecher & Johnson, 2019). The results indicated that bilingual infants outperformed monolingual infants in recognizing talkers speaking in a foreign language. Fecher and Johnson (2019) proposed multiple explanations, including enhanced processing of acoustic-phonetic detail and memory advantages in bilingual infants. However, since this study investigated a cohort of bilingual infants

who varied in the second language that they were exposed to other than English, it is difficult to establish whether bilinguals display a talker identification advantage simply because of general cognitive advantages, or whether specific mechanisms such as familiarity with phonological structure may also facilitate talker identification in bilinguals. Furthermore, it is of interest to examine how specifically bilingual children, rather than infants, perform compared to monolinguals in a similar identification task, since language exposure and cognitive mechanisms associated with language use change as a function of development.

Levi (2018) examined whether the same bilingual advantages are present in school-age bilingual children when a bilingual speaker hears a talker speaking a language that is unfamiliar to them. Levi (2018) implemented both a talker discrimination task in which the to-be-discriminated talkers spoke in either German-accented English, or German, as well as a talker-voice learning task. Even though none of the bilingual subjects spoke German, they still outperformed monolinguals in discriminating talkers speaking German (the unfamiliar language) as well as foreign-accented English (Levi, 2018). Bilingual subjects also learned to identify the voices faster than monolingual subjects (Levi, 2018). To address why bilinguals were better at discriminating talkers who spoke with unfamiliar foreign accents, Levi (2018) proposed that bilinguals may have more experience and exposure with foreign-accented speech or alternatively, have better cognitive control, meaning that they are better at focusing on tasks and suppressing irrelevant information than monolinguals. Given that previous studies have shown that bilinguals demonstrate better cognitive control (Bialystok & Martin, 2004), social

processing (Fan et al., 2015), and pitch perception (Krizman et al., 2012), Levi (2018) proposed that these enhanced abilities might explain why bilinguals outperform monolinguals in discriminating between talkers speaking in an unfamiliar language as well as identifying them faster. While Levi (2018) draws similar conclusions from both the discrimination and identification tasks, it should be pointed out that performing these different tasks might draw upon different cognitive resources.

It is important to note that it is still not entirely evident whether there is a bilingual advantage for identifying a talker that speaks in a language that is unfamiliar. As Levi (2018) did not specify the age range of children used in the study, included many different types of bilinguals speaking different languages, and included children who were exposed to another language on a daily basis but were not actually bilingual, more studies implementing additional control across variables are needed to investigate this bilingual advantage.

Identifying whether a bilingual advantage exists for talker identification tasks could provide a better understanding of the candidate mechanisms involved in identifying talkers' voices. Evaluating the impact of differences in language exposure on talker recognition performance may reveal whether familiarity effects in the context of talker identification rely on speech perception mechanisms (such as phonological familiarity and pitch perception) that are specific to particular types of language experience in bilinguals or are the result of general cognitive and perceptual mechanisms that are shared by all types of bilinguals.

Aims of the Present Study

Thus far, the literature does not clearly indicate whether a general bilingual advantage exists for talker identification or whether particular language-specific experiences modulate the ability to identify talkers' voices. To address why and how language experience, and in particular bilingual experience, influences a listener's ability to identify talker's voices, the current study examined talker identification and learning in participants with specific language backgrounds. By carefully selecting subjects based on their language backgrounds and experience, the current study attempted to isolate specific mechanisms that may allow bilinguals to identify talkers.

Since several studies suggest a bilingual advantage exists for identifying talkers who speak in a familiar foreign accent (Goggin et al., 1991; Levi, 2018; Stevenage et al., 2012), a claim that remains debatable, one goal of this study was to investigate possible experience-specific mechanisms that may aid bilingual individuals in better identifying talkers. Specifically, one mechanism of interest was familiarity with specific language phonology (Orena et al. 2019; Perrachione et al., 2011; Perrachione, 2019). To investigate whether familiarity with the sound structure of a particular language aids bilingual talker identification, this study recruited monolinguals, Spanish-English bilinguals, and other-English bilinguals[1] and familiarized them with voices speaking Spanish-accented English. The choice of Spanish-English bilinguals was to isolate the possible effect of familiarity with phonological form as a mechanism for bilingual

[1] Korean-English bilinguals were originally recruited as a comparison group unfamiliar with the Spanish-accented materials. However, due to difficulty recruiting this specific group, other language-English bilinguals were included to provide a group of bilinguals not familiar with Spanish phonology.

advantage; bilinguals who speak Spanish as one of their languages and are familiar with the language-specific phonology produced by the accented speakers would presumably perform better relative to monolinguals and other language-English bilinguals who would have less familiarity with Spanish-accented speech. Presumably, all groups would have equal familiarity with English vocabulary present within the auditory stimuli.

Alternatively, bilingual language experience, or specific types of language experience more generally, might impact talker identification in a different way. Bilinguals in general and particularly those who speak tonal languages such as Mandarin may do as well as Spanish-English bilinguals in the task as a result of utilizing a different mechanism, such as better pitch perception. Investigating pitch perception as a possible mechanism is worthwhile since listeners who have experience with tonal languages demonstrate enhanced sensitivity to pitch as well as an advantage in identifying talkers compared to bilinguals who do not speak a tonal language (Xie & Myers, 2015). This study points towards superior pitch perception being a mechanism aiding certain bilinguals in talker identification. However, it remains unclear if bilingual participants use one or multiple mechanisms separately or simultaneously when tasked with identifying talkers.

Finally, however, a general bilingual advantage might exist for adult bilinguals when identifying talkers who speak in a foreign accent, regardless of whether the accented speech does or does not match their own native language phonology. This explanation relates to the numerous studies which highlight bilinguals' enhanced cognitive control

compared to monolinguals (Bialystok & Martin, 2004; Brito & Barr, 2014; Fecher & Johnson, 2019; Levi, 2018; Liu & Kager, 2017; Orena et al., 2019; Sebastián-Gallés et al., 2012). Having enhanced cognitive control may aid all bilinguals, regardless of specific language experience, to better attend to talker-specific variation and thus outperform monolinguals in talker identification tasks.

Most of the studies discussed in this literature review employ an explicit talker identification task. During the exposure phase of an explicit talker identification task, participants are asked to click on an avatar or visual image upon hearing a talker's voice as a means of explicit identification. Given that in most communicative situations, listeners are familiarized with a voice through implicit exposure in which they attenuate to the verbal content of an individual's speech rather than their identity, this study implemented an implicit talker identification task modeled by Lee and Perrachione (2022). During the exposure phase of this task, individuals were required to complete a task in which they needed to pay attention to the verbal content of the talkers' speech, rather than explicitly identify each talker. While Lee and Perrachione (2022) have demonstrated that individuals were better able to learn and identify voices they were previously exposed to through this implicit paradigm, this study aims to replicate these findings and extend the investigation to Spanish-accented speech and bilingual talker identification.

Hypothesis

We hypothesize that bilinguals who have familiarity with specific phonology reflected by a talker's voice (for example: a Spanish-English bilingual listening to a talker speaking

English with a Spanish accent), will perform better in talker identification tasks than bilinguals who are not familiar with specific phonology conveyed by a talker's voice (for example: other language-English bilingual listening to a talker speaking English with a Spanish accent). If both groups of bilinguals outperform monolinguals, then an alternative hypothesis is possible: bilinguals in general have better cognitive control or utilization of general language cues (such as pitch), which allows them to perform better in talker discrimination tasks than monolinguals, regardless of what additional knowledge of specific language structure they have.

Methods

Participants

To test whether there is a bilingual advantage and furthermore if familiarity with phonological form is a mechanism which aids bilinguals in talker identification, I recruited monolinguals, Spanish-English bilinguals, and Korean-English bilinguals. Participants were recruited and initially screened through Prolific, a research platform for online experiments (Prolific, 2014). Participants were screened for age (ranging from 18-35), an approval rating in Prolific of 80-100, absence of speech and hearing difficulties, cochlear implants, and colorblindness. Equal numbers of female and male participants were recruited in the initial sample. For bilingual participants, the experiment additionally screened participants for being bilingual (defined as "native language + one other language"), raised with two or more languages, and fluent in English and Spanish or English and Korean depending on the specific bilingual group recruited (Prolific, 2014).

Eighty participants were recruited for the monolingual group, 120 participants for the Spanish-English bilingual group, and 50 participants for the Korean-English bilingual group. Although the online platform Prolific allowed screening for specific demographics to recruit specific types of participants including bilinguals, information from the language information questionnaire and the self-proficiency language survey (described below) revealed that many participants reported different language experiences, including reporting being bilingual in languages other than Spanish or

Korean. As there was an insufficient number of Korean bilinguals, participants were recategorized as monolingual (n=64; 29 female; 35 male), Spanish-English bilingual (n=48; 16 female, 32 male), and other-bilingual (n=76; 46 female, 30 male). The criteria for participants to be considered monolingual was if they reported that their first and native language was English and indicated that they were not fluent in languages other than English (listed in the subject language questionnaire, see Appendix B). The criteria for participants to be considered as Spanish-English bilinguals was if they indicated that their first and native language was "English and another language" or "other language," and fluency was "Spanish," and/or had a value of 1 or 2 from the self-proficiency language survey for both Spanish and English (Su & Styles, 2017; see Appendix C). The criteria for participants to be considered as other (non-Spanish) bilinguals was if they indicated that their first and native language was "English and another language" or "other language," and/or that they were fluent in any other language other than Spanish.

Stimulus Materials

Auditory Stimuli

Audio stimuli were drawn from a database of utterances collected and maintained by the Speech and Language Perception Laboratory, and used previously in other studies (e.g, Alexander & Nygaard, 2019). The database consists of recordings of six male and six female native Spanish speakers from Mexico City living in Atlanta. The speakers recorded monosyllabic words in English, 72 of which were classified as "easy", and 72 of which were classified as "hard." "Easy" versus "hard" words were chosen from a

computerized lexicon created by Luce and Pisoni (1998), which assessed word frequencies in English and their degree of phonetic similarity with other words. “Easy” words are high frequency words with low phonetic similarity with other words and “hard” words are low frequency words with high phonetic similarity with other words. The talkers were recorded in a sound-attenuated room with a SONY Digital Audio Tape Recorder TCD-D7. These recordings were re-digitized and edited using Sound Studio software (Felt Tip, Inc). Only the male talkers were used for the current study. The mean age of the six male talkers was 32.75 (ranging from 26-39) and their mean age of arrival to the US was 26.42 years (ranging from 21-34). The mean age for when the talkers started speaking English was 17 (range 2-28).

Each phase of the experiment, exposure and test, required three talkers. To create the two groups of three talkers, mean intelligibility of the talkers in each group was equated. Intelligibility of each of the speakers was assessed previously (10 participants per speaker transcribed all 144 recorded words; see Sidaras et al., 2009). Average intelligibility scores were 54.67 (SD = 5.55) for one group and 51.33 (SD = 7.80) for the other group. Of the 72 “easy” words recorded, 10 distinct words were selected for the exposure task and the test task respectively (see Appendix A). Frequency of occurrence scores for each word was assessed using the MRC Psycholinguistic database (Coltheart, 1981). The set of words for the exposure phase had an average frequency of 154.1 (SD = 104.11). The set of words for the test phase had an average frequency of 138.4 (SD = 112.77). Audio recordings were edited in Praat (Boersma & Weenink, 1991) to reduce the amplitude of clicks that occurred during recording.

Because pilot testing showed that participants performed with high overall average scores for talker identification learning during the test phase, babble noise was added to the auditory word files used in the test phase using Audacity, an audio editing and recording platform (Audacity, 2023). Discontinuous speech of multiple talkers was used as a background masking signal for auditory stimuli in the test phase as several studies have shown that while individuals are still able to find a main auditory target intelligible, their performance in tasks which require them to attend to an auditory target declines in the presence of babble background noise (Renz et al., 2018). Original babble noise audio files were created by Dr. Kate Revill. Two recordings of non-accented control talkers producing two sentences collected by the Speech and Language Perception Laboratory (Sidaras et al., 2009) were digitally combined. The content of the sentences spoken by the recorded talkers were from the Harvard Sentences set (IEEE, 1969). For this study, babble noise files were combined using Audacity with the original auditory stimuli created for the test phase and amplitude normalized to 68dB on Praat (Boersma & Weenink, 1991). Signal to noise ratios for the auditory stimuli were approximately -2dB.

Language questionnaires

Two questionnaires were used to screen participants and sort them into participant type categories. All participants (both monolingual and bilingual) were administered a language information questionnaire created by the Speech and Language Perception Laboratory (see Appendix B), in which they are asked to fill out personal information (sex, year and place of birth, hearing/speech disorders, handedness, etc.), as well as information regarding their language history (what their native language is,

whether they think they have a foreign accent while speaking English, the ages they began to use other languages, and the frequency and contexts in which they use languages, other language experience, etc). In addition to the language information questionnaire, bilingual participants were asked to complete a self-proficiency language survey (see Appendix C), in which they selected one of six statements that best described their proficiency in both English and another language (either Spanish or Korean depending on the type of bilinguals recruited; Su & Styles, 2017). Statements ranged from "I can speak easily about most things without thinking about how to say them," to "I have taken lessons and have begun practicing how to say certain kinds of words and sentences" (Su & Styles, 2017).

Study Design

Many talker identification tasks used in previous studies do not model real-life social interactions since these experiments often require participants to complete an explicit talker identification task in which they actively assign an identity to a voice. In many real-world communicative situations, however, listeners are exposed to an individual talker's identity implicitly by focusing on the verbal content of their speech. To better model situations encountered when learning voices in the real world, my study design uses an implicit talker familiarization procedure developed by Lee and Perrachione (2022), which demonstrated that listeners could learn a set of talkers' voices through a verbal 1-back task. In the current task as well, participants made judgments about the content of the speech rather than explicitly attending to the talker's voice during an exposure phase.

This study implements a modified version of Lee and Perrachione's verbal 1-back task. Words spoken and recorded by six Spanish-accented talkers were used to further simulate a real-world interaction, where a listener relies on verbal context to collect implicit information on a talker's identity. Due to the availability of stimuli, three talkers each were used for the exposure and test phases of the experiment. Spanish-accented talkers producing English words were used to specifically examine aspects of the language familiarity phenomenon.

The experiment included both exposure and test phases. In the exposure phase, listeners were presented with three different male talkers each producing one word of five-word sequences, which changed in order on every trial. Participants were asked to indicate whether the middle word (3 of 5) that they heard was the same or different as on the previous trial. This task does not require participants to explicitly attend to the talkers' voice characteristics but rather asks them to focus on the speech content (the spoken order of word).

In the test phase, subjects heard three talkers producing words presented in random order and were asked to learn to associate an avatar with each voice. Across conditions, subjects either heard the same voices as presented during exposure (familiar) or a set of novel voices (unfamiliar) that they were not previously exposed to. Better performance for voices heard during exposure (familiar) was used as an index of implicit talker learning. Lee and Perrachione (2022) found that participants were more accurate when learning to identify voices at test that they had heard previously during the exposure

phase than novel voices which they had not been exposed to, suggesting that listeners benefitted from hearing the voices presented during the exposure phase.

Procedure

The experiment was created on Gorilla, an online platform used by researchers for creating behavioral experiments (Gorilla, 2023). Participants were first required to sign and complete a consent form. Next, all participants were administered a language information questionnaire. Bilingual participants were additionally administered a self-proficiency language survey. Participants then completed a short audio test to ensure that their headphones and sound from their computer was working (Woods et al., 2017).

For the exposure phase, three avatars appeared on the screen as participants heard one of three talkers saying a series of five words. On the following trial, participants heard a talker (whether it was the same or different talker out of the three available talkers was randomized), saying another series of five words, and were then asked to indicate whether the third word in that sequence was the same or different word as heard in the previous trial. For example, a participant should click the button marking “same” if they hear “cat dog spoon loft bliss” on the previous trial and then hear “blog mess spoon jog mat” on the current trial. The order in which the three avatars were presented randomly alternated after every trial and did not serve any experimental purpose (Lee & Perrachione, 2022). The exposure task consisted of 120 trials in which each talker was featured for 40 trials each in random order. Each trial began with a fixation cross which

lasted for 300 milliseconds. Participants received feedback, “correct” or “incorrect.” after each trial. One out of every five trials had “same” as the correct response. Trials in which the correct response was “same,” were pseudo randomly presented.

For the test phase, participants viewed three new avatars presented on the screen as they heard one of three talkers say a series of five words. Unlike the exposure phase, the order in which the three avatars were presented on the screen did not change after each trial. On each trial, participants were tasked with learning which talker was associated with which avatar by clicking on the avatar they thought was associated with the voice they heard speaking the five words. The test phase consisted of 45 trials in which each talker was featured for 15 trials. Trials were presented in random order. Participants received corrective feedback after each trial. Depending on the condition, the three talkers in the test phase were either the same talkers as presented during the previous exposure phase (familiar), or different (unfamiliar).

The “familiar” and “unfamiliar” conditions were counterbalanced such that participants could be exposed and tested on either set of three talkers. For example, Condition 1 (familiar) featured talkers “SM2, SM5, and SM6” in the exposure phase, and the same talkers, “SM2, SM5, and SM6” in the test phase. Condition 2 (unfamiliar) featured talkers “SM2, SM5, and SM6” in the exposure phase, and different talkers, “SM3, SM4, and SM7” in the test phase. Condition 3 (familiar-counterbalance) featured talkers “SM3, SM4, and SM7” in the exposure phase, and talkers, “SM3, SM4, and SM7” in the

test phase. Condition 4 (unfamiliar-counterbalance) featured talkers “SM3, SM4, and SM7” in the exposure phase, and talkers, “SM2, SM5, and SM6” in the test phase.

The experiment took approximately 30 minutes for participants to complete.

Results

Exposure Phase

Performance on the exposure phase for each participant was scored as the mean accuracy of trials answered correctly out of 119 trials (the first trial was excluded from scoring given that participants did not submit a response for it). Each trial was scored dichotomously, in which a value of 0 represented an incorrect response and a value of 1 represented a correct response.

Because the exposure phase was meant to familiarize the participants with the talkers' voices, we excluded participants who scored less than 65% on the exposure phase from subsequent analysis. Using this exposure phase performance threshold, we excluded a total of 16 participants from the monolingual group (n = 64), 22 participants from the Spanish-English bilingual group (n = 48), and 25 participants from the other bilingual group (n = 75).

To determine whether participant groups differed with respect to exposure phase performance after implementing the exclusion criteria, we conducted a one-way analysis of variance (ANOVA) with exposure phase score as the dependent variable and participant type (monolinguals, Spanish-English bilinguals, and other bilinguals) as the between-subject factor. For all ANOVA tests, we used IBM SPSS Statistics (Version 29.0.2.0 (20)) with a predetermined significance level of $p = 0.05$. Among

participant types, which included monolinguals (M_1 = 0.91, SD_1 = 0.08), Spanish-English bilinguals (M_2 = 0.88, SD_2 = 0.10), and other bilinguals (M_3 = 0.90, SD_3 = 0.08), no statistically significant difference was found for exposure phase performance score [$F(2, 185) = 2.45$, $p = 0.09$; Figure 1). Given that there was no significant effect of participant type on exposure phase performance, we did not include exposure phase performance as a variable in subsequent analyses. Therefore, the same participants and participant type groupings were used in subsequent analyses examining test phase performance.

Test Phase

Performance on the test phase was scored in two ways: as an overall performance score (calculated as the mean accuracy of trials answered correctly out of the total 45 trials where a value of 0 represented an incorrect response and a value of 1 represented a correct response) and as a function of block. Performance as a function of block was measured in addition to total test phase performance in order to better understand the impact of prior exposure on each participants' learning over time. Test phase performance as a function of block was divided into five blocks, each composed of nine trials. Block performance score for the test phase was calculated as the mean accuracy of trials answered correctly for each block.

To identify possible main effects and interactions between test block performance, participant type, and experimental condition (whether participants were tested on previously exposed voices during the exposure phase- "familiar," versus novel voices-

"unfamiliar"), we conducted a three-way ANOVA. *Block* (1-5) served as the within-subject factor, and *Participant type* (1 = "monolingual," 2 = "spanish-english" bilingual, 3 = "other bilingual") and *Condition* (1 = "familiar," 2 = "unfamiliar") were the between-subject factors. We observed a main effect for *block* scores, suggesting that the overall level of performance improved across blocks [$F(4, 183) = 47.04$, $p < 0.001$; see Figures 2a-c]. A main effect for *Condition* was also observed, [$F(1,186)=5.02$, $p=0.03$]. As expected, participants performed better in the test phase when tested on voices they were previously exposed to (the "familiar" condition), versus on novel voices without previous exposure (the "unfamiliar" condition).

A main effect was also observed for *Participant Type* [$F(2, 185) = 5.00$, $p = 0.01$], which suggests that regardless of the experiment condition, participant group types differed in overall test performance. Performance for the English monolingual and other bilingual participants was better than for the Spanish bilingual participants, suggesting that language experience affected sensitivity to talker-specific information in the test phase. English monolinguals exhibited the highest performance for the test phase ($M_1 = 0.77$, $SD_1 = 0.15$), followed by other-bilinguals ($M_2 = 0.73$, $SD_2 = 0.17$). Spanish-English bilinguals scored the lowest ($M_3 = 0.68$, $SD_3 = 0.21$). Post hoc comparisons using Tukey's HSD, revealed a significant difference between English monolinguals and Spanish-English bilinguals (Tukey's HSD: $p = 0.01$). No significant differences were found between English monolinguals and other bilinguals (Tukey's HSD: $p = 0.292$) and between Spanish-English bilinguals and other bilinguals (Tukey's HSD: $p = 0.18$). These

results suggest that Spanish-English bilinguals performed significantly worse than the other two participant types overall.

A marginal *Participant type* × *Condition* interaction [$F(2, 185) = 2.54$, $p = 0.08$] was also observed, suggesting that the effects of prior exposure or familiarity may have differed across participant groups. When tested on familiar voices, English monolinguals scored the highest ($M_1 = 0.80$, $SD_1 = 0.16$), followed by Spanish-English bilinguals ($M_2 = 0.74$, $SD_2 = 0.19$), followed by other-bilinguals who scored slightly lower ($M_3 = 0.73$, $SD_3 = 0.18$). When tested on unfamiliar voices, English monolinguals and other-bilinguals scored similarly, ($M_1 = 0.74$, $SD_1 = 0.15$; $M_2 = 0.74$, $SD_2 = 0.15$) while Spanish-English bilinguals scored the lowest ($M_3 = 0.61$, $SD_3 = 0.23$). Finally, a significant interaction was observed for the *Participant type* × *Block* interaction [$F(2, 185) = 47.04$, $p = 0.01$], suggesting that the degree or rate of learning differed across participant group. No other significant interactions were observed.

Separate follow-up ANOVAs with *Condition* (familiar, unfamiliar) and *block* (1-5) as factors were run for each participant type group to examine the differences in performance as a function of these factors. A significant main effect of *Condition* was only observed for Spanish-English bilinguals [$F(1, 47) = 4.76$, $p = 0.03$] and not for monolinguals or other-bilinguals, suggesting that only Spanish-English bilinguals benefitted from familiarity to previously exposed voices (Figure 2b). Across participant types, there was a significant main effect of *block*: monolinguals [$F(4, 60) = 7.19$, $p < 0.001$], Spanish English bilinguals [$F(4, 44) = 21.65$, $p < 0.001$], and other bilinguals

[$F(4, 71) = 24.70$, $p < 0.001$]. The interaction of *Condition* and *block* did not reach significance for any participant type. Although these results confirm that each respective participant group type improved in their ability to identify talkers' voices as a function of block throughout the test phase, further analyses were performed to investigate whether participant types were learning differently from one another.

Separate one-way ANOVA analyses were performed using three separate dependent variables to quantify learning. *Participant type* was the fixed factor for all three analyses. In the first of the three analyses, degree of *learning* (quantified as the mean of each participant types' block 5 score minus the mean of each participant types' block 1 score) was the dependent variable. A significant effect of participant type for test block *learning* was found [$F(2, 185) = 4.84$, $p = 0.01$], suggesting that participant types exhibited different degrees of learning (Figure 3). Indeed, English monolinguals exhibited the lowest learning rate ($M_1 = 0.13$, $SD_1 = 0.26$), while Spanish English bilinguals had the highest learning rate ($M_2 = 0.27$, $SD_2 = 0.21$) and the other bilinguals had the second highest learning rate ($M_3 = 0.23$, $SD_3 = 0.27$). Post hoc comparisons using Tukey's HSD, revealed significant differences between English monolinguals and Spanish-English bilinguals (Tukey's HSD: $p = 0.01$) and between English monolinguals and other bilinguals (Tukey's HSD: $p = 0.04$). No significant difference was found between the Spanish English bilinguals and the other bilinguals (Tukey's HSD: $p = 0.78$), which suggests that bilinguals in general learned differently than monolinguals.

To assess whether baseline performance at test differed across participant types, the second analysis used the mean score for *block 1* as the dependent variable. Significant differences were observed in *block 1* scores across participant types [$F(2, 185) = 6.32$, $p < 0.002$; Figure 4]. English monolinguals had the highest accuracy ($M_1 = 0.70$, $SD_1 = 0.22$), followed by other bilinguals ($M_3 = 0.60$, $SD_3 = 0.25$). In contrast to our hypothesis, Spanish English bilinguals had the lowest accuracy at baseline ($M_2 = 0.55$, $SD_2 = 0.23$). Follow up comparisons revealed significant differences in performance between monolinguals and Spanish-English bilinguals (Tukey's HSD: p = 0.002) and between English monolinguals and other bilinguals (Tukey's HSD: p = 0.04). No significant differences were observed between Spanish-English bilinguals and other bilinguals (Tukey's HSD: p = 0.38), suggesting that monolingual English speakers were better able to identify talkers' voices initially.

In the third analysis, the average mean score for *block 5* served as the dependent variable. All participant types exhibited similar performance ($M_1 = 0.83$, $SD_1 = 0.20$, $M_2 = 0.81$, $SD_2 = 0.23$, $M_3 = 0.84$, $SD_1 = 0.18$) and no significant difference was observed for performance on *block 5* across participant types [$F(2, 185) = 0.27$, $p = 0.76$; Figure 5]. Taken together, these findings suggest that although performance differed at the beginning of the test phase, bilinguals learned quickly and performance was comparable at the end of the testing session.

Discussion

This study examined whether a general bilingual advantage exists for identifying talkers speaking in a familiar or unfamiliar accent. Through recruitment of specifically Spanish and non-Spanish bilingual participants as well as monolingual participants, the goal was to identify whether familiarity with phonology is a specific mechanism aiding bilinguals in identifying talkers, or whether experience with more than one language, bilingualism, confers a general advantage in the identification of talkers' voices. While all participant types engaged in talker learning, suggesting that listeners learned aspects of foreign-accented talkers' voices through implicit exposure (Lee & Perrachione, 2022), we observed differences in performance across participant types. Overall, monolinguals were better at learning to identify talkers at test (regardless of voice familiarity) than either bilingual participant group, suggesting that a general bilingual advantage may not exist. It appeared that only Spanish-English bilinguals benefited from previous exposure to voices as they performed significantly better when tested on familiar voices. Such findings indicate that accent and phonology-related familiarity may be a mechanism at play during talker identification tasks.

Overall Performance Across Participant Types

In terms of overall performance during the test phase, monolinguals outperformed both Spanish-English bilinguals and other-bilinguals. We conclude therefore, that there does not appear to be a bilingual advantage for listeners when identifying talkers speaking in a foreign accent. Additionally, when tasked with identifying novel voices during the test phase, Spanish-English bilinguals exhibited the lowest overall performance. These

findings were contrary to our predictions. We hypothesized that the Spanish-English bilinguals would outperform both the other-bilingual group and monolinguals given their familiarity with specific phonology reflected by the talkers' voices.

One possible explanation for why monolinguals outperformed both bilingual participant groups, is that they were more familiar with the vocabulary of the English words spoken by the talkers and thus benefitted from the language familiarity effect (Perrachione, 2019). However, this would be surprising as these words were all classified as "easy" high frequency words (Luce & Pisoni, 1998). While we ensured that all bilinguals were proficient in English, our questions within the language information questionnaire focused on participants' experience with non-English languages (such as age of acquisition, how often they speak in said language, etc) rather than on their English proficiency. In the future, it would be of interest to collect more information pertaining to bilinguals' experience with English to ensure that their familiarity with vocabulary was comparable to monolinguals'.

There are many possible explanations as to why these Spanish-English bilingual participants did not demonstrate an advantage as predicted when identifying talker's voices speaking in a Spanish accent. The first being that although participants were fluent in Spanish, the origin of their Spanish did not align with the particular Mexican-Spanish accent reflected in the auditory speech stimuli (Alexander & Nygaard, 2019) that were used. It remains unclear whether listeners that speak the same language, but with different regional and national accents, would benefit from language familiarity in

talker identification (Stevenage et al., 2012). To infer what type of Spanish accent Spanish-English bilingual participants were familiar with, we examined "place of birth" responses collected through the language information questionnaire. We found that most participants reported their place of birth as either the United States or a specific state (n=44) while several were born in other countries such as the Dominican Republic (n=2), Peru (n=1) and Nigeria (n=1). Of the 28 participants that reported a specific state within the U.S. as their place of birth, 15 reported being born in either California or Texas, the states with the highest levels of Mexican immigrants (Migration Policy Institute, 2022). In the future, we hope to collect more specific information regarding familiarity with specific types of Spanish accents as well as family history information pertaining to nationality and ethnicity.

Another explanation as to why Spanish-English bilinguals performed significantly worse than monolinguals is the possible relationship between exposure and test phase performance. Although no significant differences in exposure phase performance was found across participant type groups, it is numerically evident that Spanish-English bilinguals scored the lowest during the exposure phase. Participants' performance in the exposure phase may therefore have affected test performance. Indeed, overall exposure score performance was significantly correlated with overall test score performance for all participant types [English monolinguals: $r(64) = 0.42$, $p < 0.001$; Spanish-English bilinguals: $r(48) = 0.52$, $p < 0.001$; other-bilinguals: $r(75) = 0.42$, $p < 0.001$; see Figure 1]. In a follow up exploratory analysis, we further examined the effect of exposure performance by limiting participants to those who scored higher than 80% rather than

65% (the original baseline) on the exposure phase. Comparing participants' performance during the test phase as a function of block, we found a similar pattern of results when looking at this new subset of participants with an 80% cutoff for exposure phase performance. However, we do see one difference between the two analyses such that a main effect of participant type on test performance was only observed for participants who scored 65% and above and not for those that scored 80% and above, suggesting that test phase performance was dependent on exposure phase performance. Given that the Spanish-English bilingual participants performed lower on average during the exposure phase performance, it is therefore possible that their exposure phase performance influenced their test phase performance. In the future, we hope to recruit more participants to better define this exposure-test relationship.

Learning Differences Across Participant Types

Although monolinguals outperformed both bilingual groups in the test phase overall, when comparing performance in the last block versus the first block of the test phase, we found that both Spanish-English and other-bilinguals learned faster than monolinguals who exhibited the lowest learning rate. Since Spanish-English bilinguals had the highest learning rate, we conclude that the Spanish-English bilinguals demonstrated a robust learning effect. Given that initial performance during block 1 was significantly lower for both Spanish-English and other-bilinguals than monolinguals, there may be a possible ceiling effect such that learning cannot be accurately measured for monolinguals.

Although speculative, a hierarchy of familiarity effects could provide an explanation as to why Spanish-English bilinguals demonstrated a robust learning effect and specifically benefitted from previous exposure to voices. Investigating both talker and language familiarity effects within a voice identity sorting task, Njie et al. (2023) recruited four listener groups who were either familiar or unfamiliar with the TV show *Derry Girls* (reflecting talker identity familiarity) and were either familiar or unfamiliar with a Northern Irish accent which was spoken by the *Derry Girls* characters (reflecting accent familiarity). While both types of familiarity benefitted participants' performance on a voice identity sorting task, the effects of accent familiarity were overall smaller and more variable than the effects of talker familiarity, suggesting a one-way dependency of accent familiarity on talker familiarity. This one-way dependency and larger effect of both accent and talker familiarity as opposed to just accent familiarity, may support why Spanish-English bilinguals, the only participant group assumed to be familiar with the talkers' accents in our study, learned to identify talkers the fastest. Additionally, this hierarchical effect would only be effective for Spanish-English bilingual participants in the familiar condition in which participants are tested on voices that they were previously exposed to, considering that the Spanish-English bilinguals were the only participant group that tended to benefit from the familiar condition as opposed to the unfamiliar. Even though Spanish-English bilinguals may still have been familiar with the auditory speech stimuli's accents when tested on novel voices, it is possible that the effects of accent familiarity on its own without familiarity of a talker's identity, are too minimal to provide a significant benefit to a listener when identifying an unfamiliar talker speaking in a familiar accent. Rather than exhibiting a general bilingual

advantage arising from factors such as better cognitive control, bilinguals of specific language experience may be utilizing accent and phonology-related familiarity as a mechanism when identifying familiar talkers.

A Bilingual Disadvantage

While our study was designed to investigate a possible bilingual advantage, it is worthwhile to consider the possibility of a bilingual disadvantage. Strong evidence suggests that even when listening to talkers speaking in a familiar language, bilinguals perform significantly worse during speech perception tasks under noisy and adverse conditions than monolinguals; in quiet conditions however, both bilinguals and monolinguals perform similarly (Mayo et al., 1997; Rogers et al., 2006; von Hapsburg & Bahng, 2009; Weiss & Dempsey, 2008). Reasons for why a noisy environment may impair performance specifically for bilinguals as compared to monolinguals during speech perception tasks include the degree of exposure and fluency a bilingual has to a language as opposed to a monolingual (Florentine, 1985), the age of second language acquisition for bilinguals and whether they learned another language within a sensitive period of development (Florentine, 1985; Mayo et al., 1997), or that bilingualism in general impacts language processing differently than for monolinguals (Weiss & Dempsey, 2008).

Although our study involved a talker identification task as opposed to a speech perception task, which requires listeners to attenuate to talkers' speech intelligibility rather than their identity, these studies examining speech processing in noisy environments are still applicable since at test babble background noise was included for

our auditory speech stimuli. For example, one specific study conducted by Tabri et al. (2011), which used English monosyllabic nouns as target words and included background babble noise in their auditory speech stimuli, found that while both bilinguals and trilinguals performed similarly on a speech perception task to monolingual participants in quiet conditions, performance declined significantly for the bi- and trilingual participants relative to monolinguals in noisy conditions. Since all participants learned their respective languages prior to the age of six, Tabri et al. (2011) proposed that these observed speech perception differences may be a result of perceptual challenges that all bilinguals experience, rather than being limited to bilinguals who learned a second language after a sensitive period in development (Florentine, 1985). von Hapsburg and Peña (2002) suggest that under noisy conditions, bilinguals experience a cost in perception as they may be spending more time searching lexicons or accessing phonemes of both of their native languages even when listening to only one familiar language. Out of all bilingual participants in the current study (n=123), 84 reported learning their first language within the age range of 0-3 years, 27 between the age range of 4-10 years, six between the age range of 11-14 years, three reported learning later than 18 years of age, and three reported "other." Since most participants reported learning their second language within the sensitive period threshold (Florentine, 1985), it is possible that the Spanish-English bilinguals performed more poorly in the test phase, even when the talkers' accent was familiar to them, because they spent more time accessing lexicons and phonemes pertaining to both Spanish and English. A similar explanation can be applied to the other-bilingual participant group; it is possible that hearing talkers speak in a foreign accent under noisy conditions prompts

bilinguals in general to access aspects of language structure, such as vocabulary and phonology, with respect to both languages which they are proficient in. The monolinguals may have exhibited enhanced performance because they necessarily only accessed their English lexicon and phonetic knowledge when listening to the foreign-accented speech. In the future we hope to create a study design that allows us to compare the effects of both the presence and absence of babble background noise within auditory speech stimuli to discern if a bilingual disadvantage is still prevalent across listening contexts.

Voice Familiarity Paradigm

This study adapted an implicit talker exposure task (Lee & Perrachione, 2022) in order to determine if talker voices could be learned implicitly and if that learning differed across listeners with different types of language experience. Applying this implicit talker learning paradigm with both monolinguals and bilinguals, we found that all participants learned to identify voices at test and in general, all participants were better at identifying voices which they were previously exposed to, as opposed to voices that were novel, although this finding was conditioned by follow-up analyses suggesting that Spanish-English bilinguals were differentially sensitive to prior exposure. As mentioned previously, our findings provide evidence for implicit learning as a way for listeners to identify talkers' voices. However, just as in Lee and Perrachione (2022), avatars were presented on the screen during both the exposure and test phase. Although no information was explicitly disclosed to participants during exposure indicating that there were three different talkers, it is possible that simply by visualizing three avatars

during the exposure phase of the experiment, participants were able to conclude that they had to discriminate between three voices. Knowing the number of talkers may have made the talker identification task easier, allowing bilinguals to perhaps rely less on specific speech perception mechanisms. Several studies investigating voice perception and familiarity effects, such as Njie et al. (2023), implement a clustering method as a way of measuring voice identification as well as accent and talker familiarity. Upon exposure to auditory speech stimuli, this task requires participants to sort voice recordings into clusters categorized by perceived talker identity without knowing the true number of different talkers. In the future, we hope to implement this method to discern whether clustering yields the same effects as observed through the presentation of avatars.

Conclusion

In this study, we were interested in how language experience influences listeners' ability to identify individual talkers. In particular, we investigated whether bilinguals demonstrate an advantage over monolinguals in identifying talkers speaking in a foreign accent which was or was not familiar. In attempts to discern a specific mechanism aiding this possible bilingual advantage, we sought to identify whether bilinguals who were familiar with the phonology reflected by the Spanish-English accented talkers would outperform bilinguals who were unfamiliar with such phonology. We predicted that the Spanish-English bilinguals who were presumably familiar with the talkers' Spanish accent would outperform non-Spanish-English bilinguals who were presumably unfamiliar with such accents. Contrary to our hypothesis, our results suggest that a general bilingual advantage in identifying talkers speaking in a foreign accent may not

exist since monolinguals outperformed both bilingual participant groups. However, since only the Spanish-English participants significantly benefited from previous exposure with the accented-talkers' voices, these findings indicate that enhanced performance in talker identification tasks may depend on an individual having familiarity with specific aspects of language structure, rather than a general bilingual advantage.

This study has several limitations including a relatively small sample size, limited information regarding subjects' familiarity with specific regional accents and nationalities, and our restricted content and usage of babble background noise within the auditory speech stimuli. Future research should expand on our investigation by comparing performance outcomes for auditory speech stimuli with and without the presence of background babble noise, recruiting more participants of different language backgrounds, and revising the task design to include a clustering method to discern whether listeners are still able to identify talkers implicitly.

As the number of bilingual children and adults who speak more than one language continues to rise globally (Byers-Heinlein & Lew-Williams, 2013), it is important to better understand the effects of language development and experience on learning and perception. Currently, additional research is needed to understand whether listening to speech produced by talkers speaking in a familiar, unfamiliar, or foreign accented language affects bilinguals and monolinguals' perception differently. We hope that our findings and discussion prompt more research in this field to better understand the

effect of listening to familiar versus unfamiliar language and accents within different social contexts. Further research will ultimately benefit educational and occupational systems in shaping curriculum, classroom, and workplace environments to optimize learning and inclusivity for students and adults of all language backgrounds.

References

Alexander, J. E. D., & Nygaard, L. C. (2019). Specificity and generalization in perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America, 145*(6), 3382. doi: 10.1121/1.5110302

Audacity [Computer software]. (2023). Retrieved from https://www.audacityteam.org.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science, 7*(3), 325–339. doi: 10.1111/j.1467-7687.2004.00351.x

Boersma, P., & Weenink, D. (1991). Praat: Doing phonetics by computer (Version X.XX) [Computer software]. Retrieved from https://www.praat.org/

Bradley & Bryant (1983). Categorizing sounds and learning to read—a causal connection. *Nature, 301*(5899), 419–421. doi: 10.1038/301419a0

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*(1), 85–95. doi: 10.1016/j.cognition.2013.09.010

Brito, N., & Barr, R. (2014). Flexible memory retrieval in bilingual 6-month-old infants. *Developmental psychobiology*, *56*(5), 1156–1163. doi: 10.1002/dev.21188

Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the Early Years: What the Science Says. *LEARNing landscapes*, *7*(1), 95–112.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(4), 497–505. doi: 10.1080/14640748108400805

Díaz, B., Cordero, G., Hoogendoorn, J., & Sebastian-Galles, N. (2022). Second-language phoneme learning positively relates to voice recognition abilities in the native language: Evidence from behavior and brain potentials. *Frontiers in Psychology, 13*, 1008963. doi: 10.3389/fpsyg.2022.1008963

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The Exposure Advantage: Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological Science, 26*(7), 1090–1097. doi: 10.1177/0956797615574699

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fecher, N., & Johnson, E.K. (2018a). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America, 143*(4), 2409–2418. doi: 10.1121/1.5032199

Fecher, N., & Johnson, E.K. (2018b). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1911–1920. doi: 10.1037/xlm0000555

Fecher, N., & Johnson, E. K. (2019). Bilingual infants excel at foreign-language talker recognition. *Developmental Science, 22*(4), e12778. doi: 10.1111/desc.12778

Flege, J E, & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America, 91*(1), 370–389. doi: 10.1121/1.402780

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

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 101–131). Lawrence Erlbaum Associates Publishers.

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

Gabrieli J. D. (2009). Dyslexia: a new synergy between education and cognitive neuroscience. *Science, 325*(5938), 280–283. doi: 10.1126/science.1171999

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & cognition*, *19*(5), 448–458. doi: 10.3758/bf03199567

Gorilla. (2023). Retrieved from https://app.gorilla.sc/login

Holmes, E., To, G., & Johnsrude, I. S. (2021). How long does it take for a voice to become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training. *Psychological Science, 32*(6), 903–915. doi: 10.1177/0956797621991137

IBM Corp. (2020). *IBM SPSS Statistics*. Retrieved from https://www.ibm.com/analytics/spss-statistics-software

IEEE. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics, 17*(3), 225-246. doi: 10.1109/TAU.1969.1162058

Kreitewolf J., Mathias S. R., von Kriegstein K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Frontiers in Psychology, 8,* Article 1584. doi: 10.3389/fpsyg.2017.01584.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences of the United States of America, 109*(20), 7877–7881. doi: 10.1073/pnas.1201575109

Lee, J. J., & Perrachione, T. K. (2022). Implicit and explicit learning in talker identification. *Attention, Perception & Psychophysics, 84*(6), 2002-2015. doi: 10.3758/s13414-022-02500-8

Levi, S. V. (2018). Another bilingual advantage? *Perception of talker-voice information. Bilingualism (Cambridge, England), 21*(3), 523-536. doi:10.1017/S1366728917000153

Liu, L., & Kager, R. (2017). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition, 20*(3), 561–575. doi: 10.1017/S1366728916000183

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and hearing, 19*(1), 1–36. doi: 10.1097/00003446-199802000-00001

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of speech, language, and hearing research: JSLHR*, *40*(3), 686–693. doi: 10.1044/jslhr.4003.686

Migration Policy Institute. (2022). Mexican Immigrants in the United States. Retrieved from https://www.migrationpolicy.org/article/mexican-immigrants-united-states#:~:text=Within%20the%20United%20States%2C%20almost,live%20in%20California%20or%20Texas.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & psychophysics*, *47*(4), 379–390. doi: 10.3758/bf03210878

Njie, S., Lavan, N., & McGettigan, C. (2023). Talker and accent familiarity yield advantages for voice identity perception: A voice sorting study. *Memory & cognition*, *51*(1), 175–187. doi: 10.3758/s13421-022-01296-0

Nygaard L. C., Pisoni D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376. doi: 10.3758/BF03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech Perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46. doi: 10.1111/j.1467-9280.1994.tb00612.x

Orena, A. J., Polka, L., & Theodore, R. M. (2019). Identifying bilingual talkers after a language switch: Language experience matters. *The Journal of the Acoustical Society of America, 145*(4), EL303. doi: 10.1121/1.5097735

Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia, 45,* 1899–1910. doi: 10.1016/j.neuropsychologia.2006.11.015

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. C. M. (2009). Differential Neural Contributions to Native- and Foreign-Language Talker Identification. *Journal of Experimental Psychology Human Perception & Performance, 35*(6):1950-60. doi: 10.1037/a0015869

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science (New York, N.Y.), 333*(6042), 595. doi: 10.1126/science.1207327

Perrachione, T. K. (2019). Recognizing speakers across languages. S. Frühholz & P. Belin (Eds.), The Oxford Handbook of Voice Perception (pp. 459-477). Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198743187.013.23

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184. doi: 10.1121/1.1906875

Prolific. (2014). Prolific: The online participant recruitment platform for research. Retrieved from https://www.prolific.co/

Renz, T., Leistner, P., & Liebl, A. (2018). Auditory distraction by speech: Can a babble masker restore working memory performance and subjective perception to baseline? *Applied Acoustics, 137,* 151-160. doi: 10.1016/j.apacoust.2018.02.023

Rogers CL, Lister JJ, Febo DM, Besing JM, Abrams HB. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(3):465–85. doi: 10.1017/S014271640606036X

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

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America, 125*(5), 3306–3316. doi: 10.1121/1.3101452

Stevenage, S. V., Clarke, G., and McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology, 24,* 647–653. doi: 10.1080/20445911.2012.675321

Su, N., & Styles, S. J. (2017). Is a high tone pointy? Speakers of different languages match Mandarin Chinese tones to visual shapes differently. *Frontiers in Psychology, 8,* Article 2139. doi: 10.3389/fpsyg.2017.02139.

Tabri, D., Abou Chacra, K. M., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International journal of language & communication disorders*, *46*(4), 411–422. doi: 10.3109/13682822.2010.519372

Thompson, J. B. (1987). Language and ideology: A framework for analysis. *The Sociological Review, 35*(3), 516–536. doi: 10.1111/j.1467-954X.1987.tb00554.x

von Hapsburg, D., & Bahng, J. (2009). Effects of noise on bilingual listeners' first language (L1) speech perception. *Perspectives on Hearing and Hearing Disorders: Research and Diagnostics, 13*(1), 21. doi: 10.1044/hhd13.1.21

von Hapsburg, D., & Peña, E. D. (2002). Understanding bilingualism and its impact on speech audiometry. *Journal of speech, language, and hearing research: JSLHR*, *45*(1), 202–213. doi: 10.1044/1092-4388(2002/015)

Weiss, D., & Dempsey, J. J. (2008). Performance of bilingual speakers on the English and Spanish versions of the Hearing in Noise Test (HINT). *Journal of the American Academy of Audiology*, *19*(1), 5–17. doi: 10.3766/jaaa.19.1.2

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America, 123*(6), 4524–4538. doi: 10.1121/1.2913046

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79*(7), 2064-2072. doi: 10.3758/s13414-017-1361-2

Xie, X., & Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America, 137*(1), 419–432. doi: 10.1121/1.4904699

Yonan C. A., Sommers M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging, 15*(1), 88–99. doi: 10.1037/0882-79

Figures

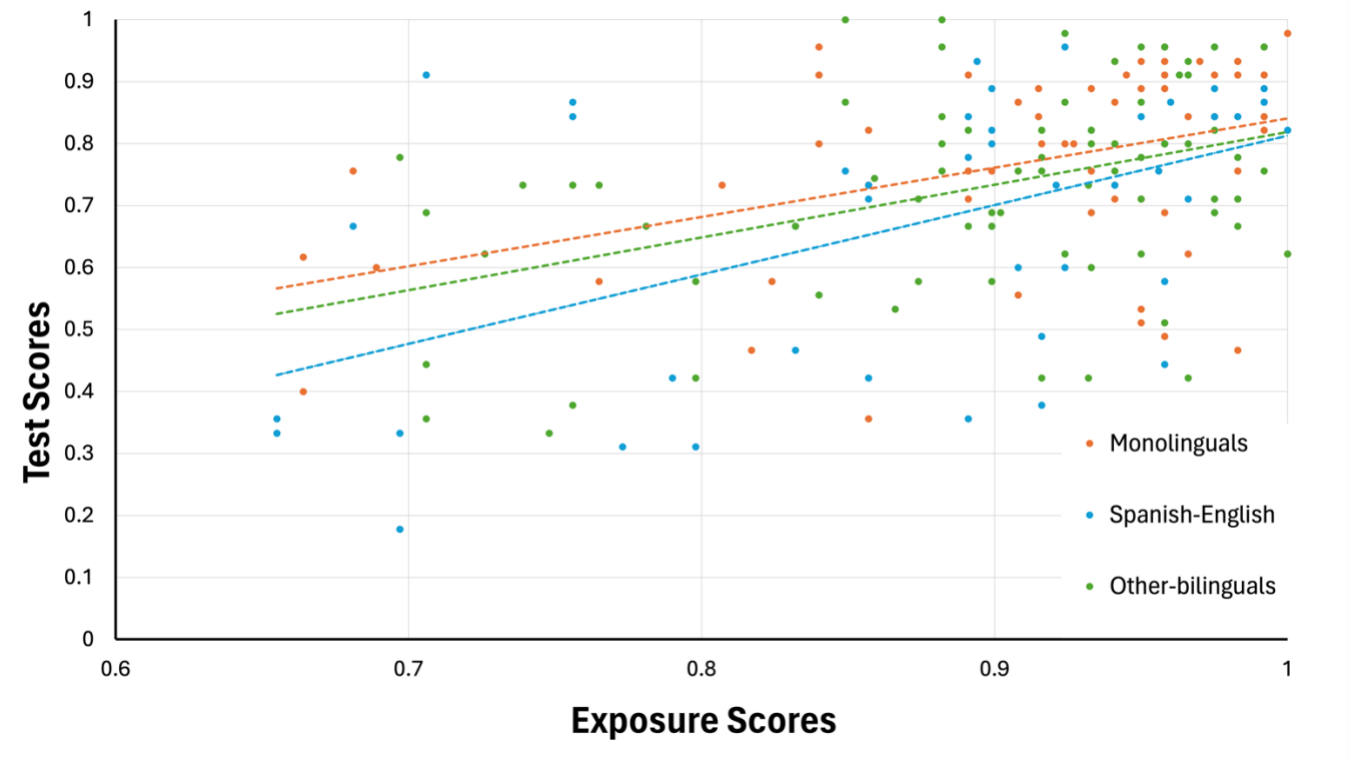
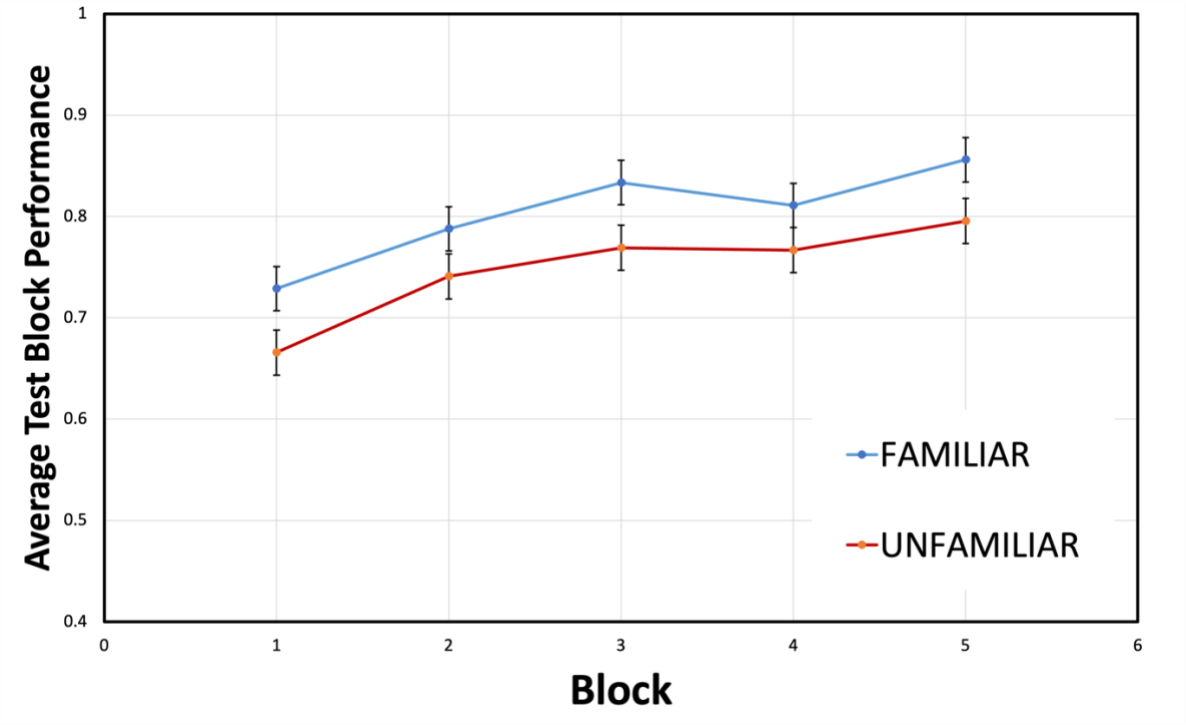


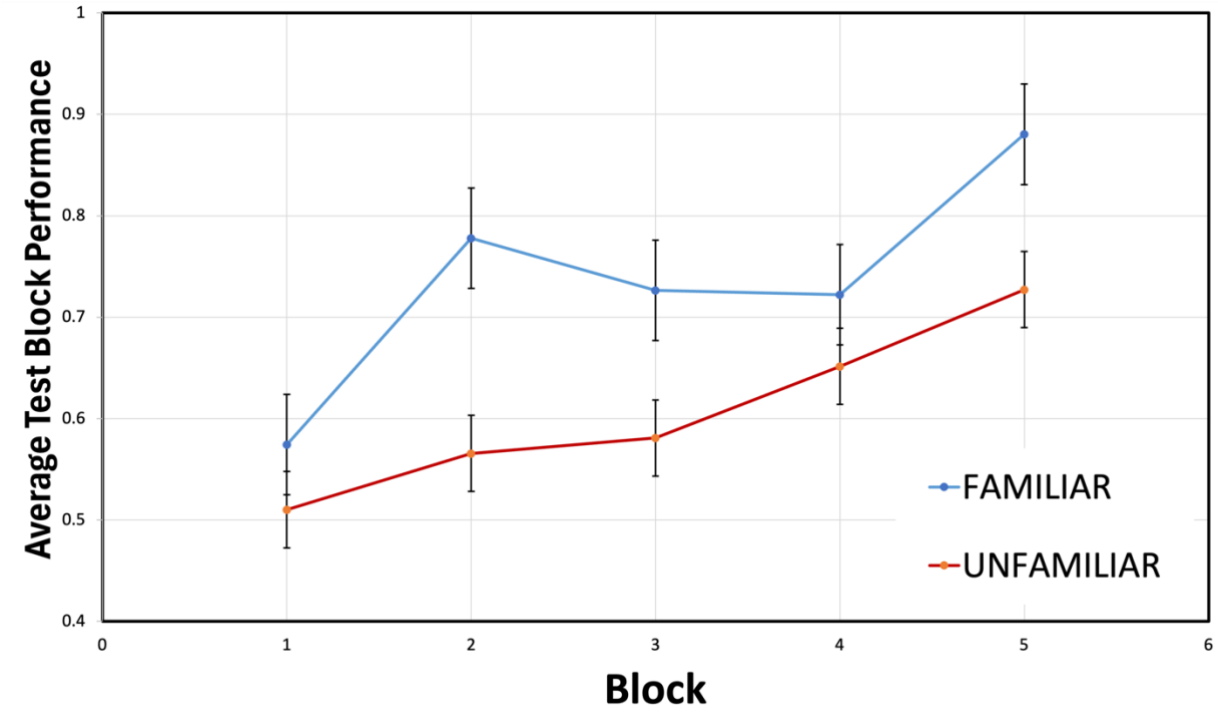
Figure 1

Mean test accuracy overall by mean exposure accuracy overall and participant type (monolingual vs. Spanish-English bilingual vs. other-bilingual)

2a. Monolinguals



2b. Spanish-English Bilinguals



2c. Other-Bilinguals

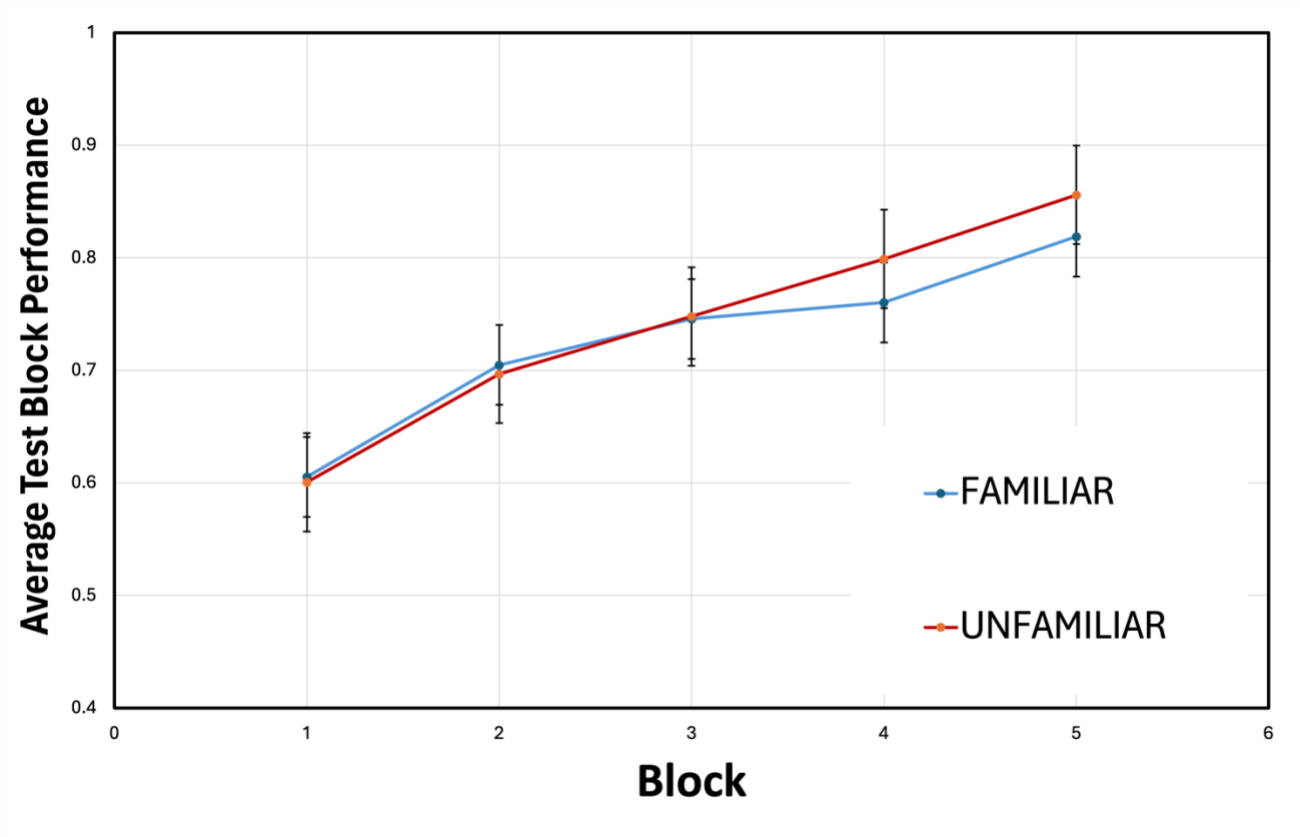


Figure 2

Mean test accuracy by block (9 trials each) as a function of condition (familiar vs. unfamiliar) for each of the three participant group types. **a** Monolinguals (n=64), **b** Spanish-English bilinguals (n=48), **c** other-bilinguals (n=75).

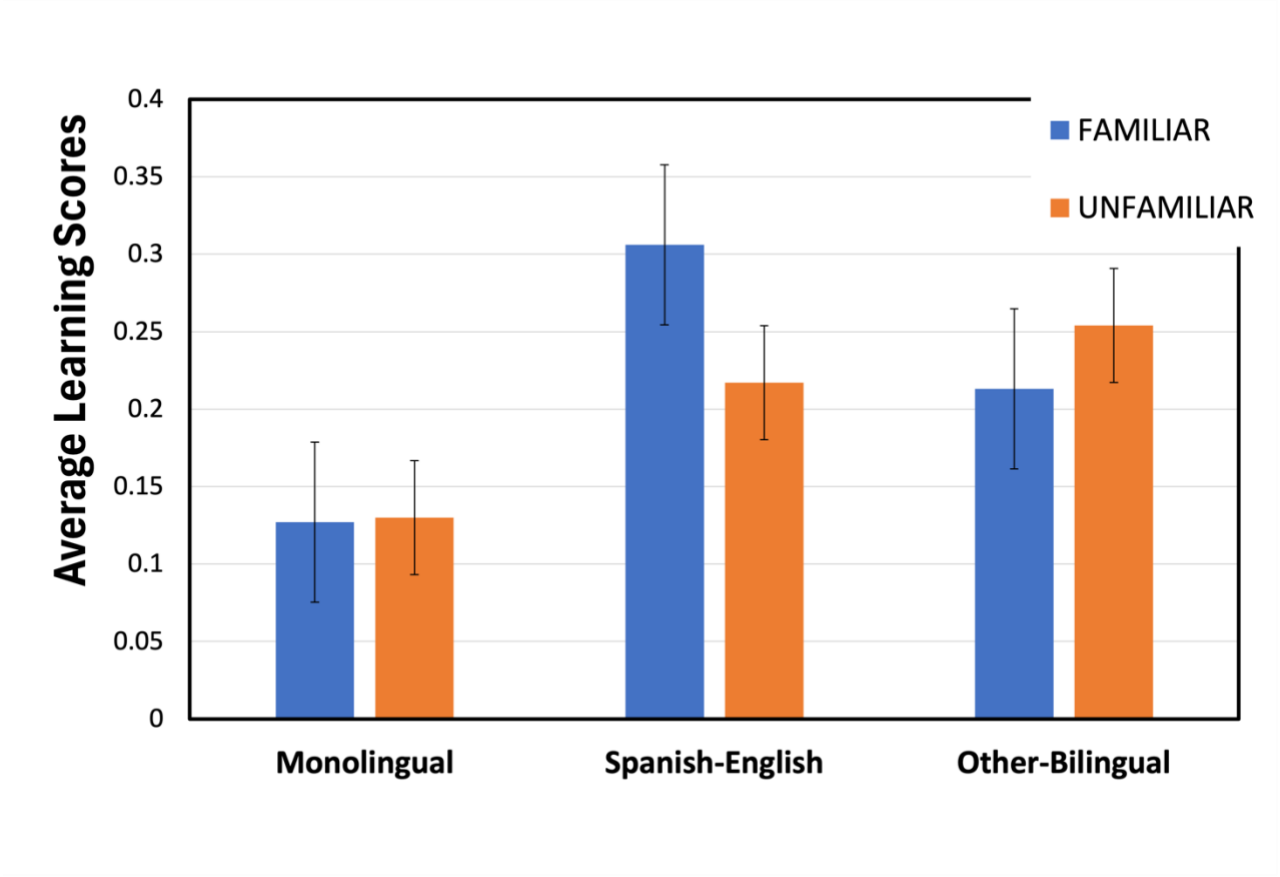


Figure 3

Mean learning scores (block 5-1) during the test phase as a function of participant group types and condition (familiar vs. unfamiliar).

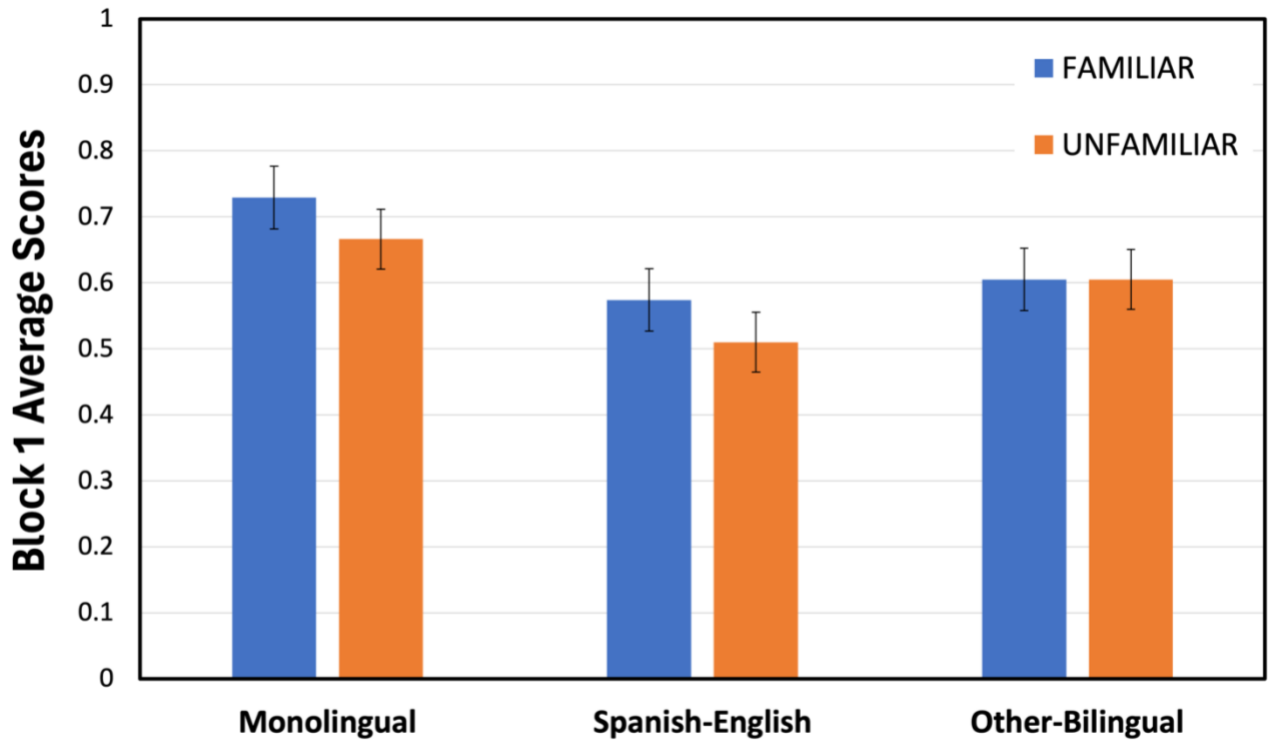


Figure 4

Mean scores for block 1 during the test phase across all participant group types and as a function of condition (familiar vs. unfamiliar).

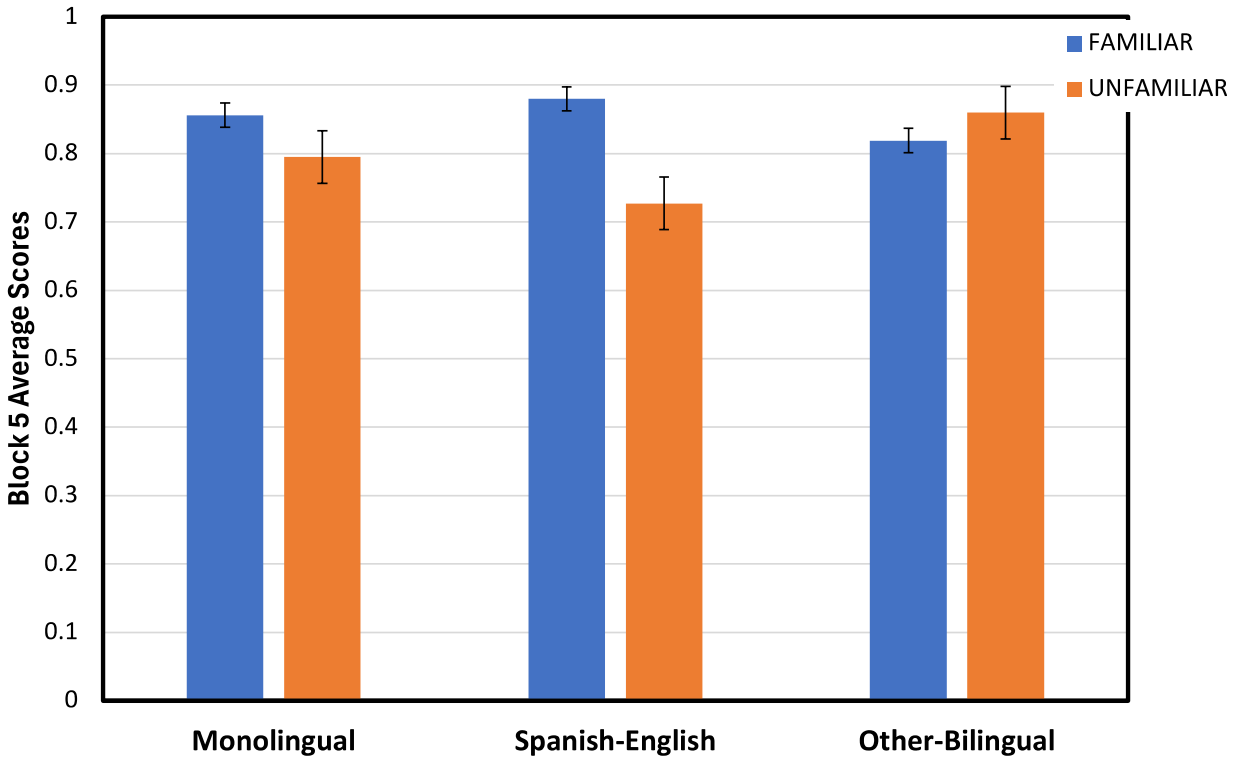


Figure 5

Mean scores for block 5 during the test phase across all participant group types as a function of condition (familiar vs. unfamiliar).

Appendix

Appendix A. Words List for Auditory Stimuli

Words selected for the exposure and test respectively were recorded by Sidaras et al. (2009)

Exposure Phase	Test Phase
dirt	cause
gave	dog
gas	fig
king	hung
page	job
pool	rough
shop	ship
thing	south
voice	theme
wife	young

Appendix

Appendix B. Language Information Questionnaire

The Language Information questionnaire was created by the Speech and Language Perception Laboratory.

1. Sex:

2. Year of Birth (yyyy):

3. Place of birth (state if USA, country if not USA):

4. What is your native language? (i.e. the language you learned to speak first)

- ◯ English
- ◯ Another language
- ◯ I learned English and another language simultaneously

First language spoken if NOT English, or language learned simultaneously with English:

5. Do you have a foreign accent when speaking English?

- ◯ No
- ◯ Yes (please explain whether you are aware of your accent or if you have been told this by other English speakers)

6. Have you ever had a hearing or speech disorder?

- ◯ No
- ◯ Yes (please explain)

7. Are you:

- ◯ Right-handed
- ◯ Left-handed
- ◯ Ambidexterous

8. Are you a fluent speaker of a foreign language or languages?

- ◯ No
- ◯ Yes (please specify)

At what age did you learn this language

- ◯ 0-3
- ◯ 4-10
- ◯ 11-14
- ◯ 15-18
- ◯ 18-25
- ◯ 26-30
- ◯ over 30
- ◯ other

How often do you still speak this language?

- ◯ daily
- ◯ weekly
- ◯ monthly
- ◯ annually
- ◯ almost never
- ◯ other

In what contexts do you most often use this language?

- ◯ social
- ◯ family
- ◯ professional
- ◯ academic
- ◯ other

What language would you say you 'think in' most often?

9. Have you studied or otherwise been exposed to other languages you are not fluent in?

- ◯ No
- ◯ Yes

If yes, please list the languages, the nature of exposure (school, caregiver, etc), and your self-assessed fluency (beginner, intermediate, conversational) in those languages below:

Appendix

Appendix C. Self-Proficiency Language Survey: Spanish-English

The self-proficiency language survey was adapted from Su & Styles (2017) and was administered to recruited bilingual participants (Spanish-English and Korean-English respectively)

Please select the statement that you feel best describes your English proficiency:

◯ I can speak easily about most things without thinking about how to say them
◯ I can talk about many things but I often need to think about how to put my words together when I speak
◯ I can understand most of what people are saying but I can only think how to reply sometimes
◯ I can understand some of what people are saying but I usually can't think how to reply
◯ I have taken lessons and have begun practicing how to say certain kinds of words and sentences
◯ None

Please select the statement that you feel best describes your Spanish proficiency:

◯ I can speak easily about most things without thinking about how to say them
◯ I can talk about many things but I often need to think about how to put my words together when I speak
◯ I can understand most of what people are saying but I can only think how to reply sometimes
◯ I can understand some of what people are saying but I usually can't think how to reply
◯ I have taken lessons and have begun practicing how to say certain kinds of words and sentences
◯ None

Appendix C. Self-Proficiency Language Survey: Korean-English

The self-proficiency language survey was adapted from Su & Styles (2017) and was administered to recruited bilingual participants (Spanish-English and Korean-English respectively)

Please select the statement that you feel best describes your English proficiency:

◯ I can speak easily about most things without thinking about how to say them
◯ I can talk about many things but I often need to think about how to put my words together when I speak
◯ I can understand most of what people are saying but I can only think how to reply sometimes
◯ I can understand some of what people are saying but I usually can't think how to reply
◯ I have taken lessons and have begun practicing how to say certain kinds of words and sentences
◯ None

Please select the statement that you feel best describes your Korean proficiency:

◯ I can speak easily about most things without thinking about how to say them
◯ I can talk about many things but I often need to think about how to put my words together when I speak
◯ I can understand most of what people are saying but I can only think how to reply sometimes
◯ I can understand some of what people are saying but I usually can't think how to reply
◯ I have taken lessons and have begun practicing how to say certain kinds of words and sentences
◯ None