2015

Lexical effects in talker identification

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Thesis

LEXICAL EFFECTS IN TALKER IDENTIFICATION

by

REBECCA LEMBER

B.A., Tufts University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2015

Approved by

First Reader _____

Tyler Perrachione, Ph.D.
Assistant Professor of Speech, Language and Hearing Sciences

Second Reader _____

Sudha Arunachalam, Ph.D.
Assistant Professor of Speech, Language and Hearing Sciences

Third Reader _____

Karole Howland, Ph.D., CCC-SLP
Clinical Assistant Professor of Speech, Language and Hearing Sciences

ACKNOWLEDGMENTS

Over the past two years, I have benefited from the help of many members of the Boston University community. I am deeply appreciative of the continuous encouragement, insight, and guidance of my advisor, Dr. Tyler Perrachione. His patient, enthusiastic explanations greatly supported my academic growth and motivation. Committee members Dr. Sudha Arunachalam and Dr. Karole Howland also generously gave their time and thoughtful feedback over the past two years. The supportive members of the Communication Neuroscience Research Lab were instrumental in the creation of this project. Much of the work presented in this thesis resulted from the collective efforts of lab members. Specifically, I would like to acknowledge Deirdre McLaughlin, for her invaluable efforts to design and build the experiment in PsychoPy, run participants, create unique word and non-word sentences, run analyses in R, and create figures; Sara Dougherty, for her substantial work to design the experiment, create non-word sentences, analyze data, and run participants; and, Ja Young Choi, for her assistance running participants and expert guidance regarding data analysis in R. Finally, I am thankful for the contributions and support of the Boston University Master's of Speech-Language Pathology students, several of whom learned and recorded the sentences used in this experiment.

LEXICAL EFFECTS IN TALKER IDENTIFICATION

REBECCA LEMBER

ABSTRACT

Adult listeners more accurately identify talkers speaking a known language than a foreign language (Thompson, 1987), a phenomenon known as the language-familiarity effect (Perrachione & Wong, 2007). Two experiments explored how knowledge of a language facilitates talker identification. In Experiment 1, participants identified talkers in three conditions: (a) a foreign-language speech condition featuring unfamiliar sound patterns and no known words; (b) a nonsense speech condition featuring all the familiar sound patterns of their native language, such as familiar phonemes, prosody, and syllable structure, but no actual words; and (c) a native-language condition with all the familiar components of a language, including words. In Experiment 2, participants again identified speakers in familiar and unfamiliar languages. In both languages, listeners identified speakers in a condition in which no word was ever repeated, and in a condition featuring repeated words. The results suggest that access to familiar, meaningful spoken words confers an advantage beyond access to familiar sounds, syllables, and prosody, particularly when words are repeated. Together, Experiments 1 and 2 support integrated models of voice and language processing systems, and indicate that access to meaningful words is a crucial component of the language-familiarity effect in talker identification.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Across speakers, there is substantial variation in the acoustic properties of speech (Hillenbrand & Houde, 2003). Differences in anatomy, physiology, and accent combine with variables such as coarticulatory effects and learned behaviors to render speech highly idiosyncratic (Munson & Babel, 2007; Skuk & Schweinberger, 2014). Talker identification, an important social skill, requires implicit processing of not only this inter-speaker variability, but also intra-speaker variability.

Adults typically excel at talker identification in their native language. There is no observed upper limit to the number of personally familiar speakers an individual can recognize with considerable accuracy (Kreiman & Siditis, 2011). However, accuracy is significantly reduced when listeners identify talkers speaking a foreign language (Thompson, 1987), a phenomenon known as the language-familiarity effect (Perrachione & Wong, 2007). It is unclear why understanding the language spoken confers an advantage for identifying talkers. Familiarity with the sounds of a language may allow listeners to perceive subtle talker variability in sound production and compare abstract internal phonological representations to the idiosyncratic pronunciations of a speaker. Similarly, linguistic knowledge of word-level differences between speakers may further support recognition of speakers and comparisons between speakers. Episodic models of lexical access state that listeners' representations of words are comprised of accumulated information about each time a particular word is spoken, including features specific to individual talkers (Goldinger, 1998). According to episodic models, listeners form representations of both abstract words and their phonetic realization. These

representations could be used to facilitate the identification of talkers' voices, particularly when they repeat the same word.

The following project explores how familiar sounds and repeated words support talker identification. Two separate experiments seek to measure and disentangle the roles of linguistic and phonetic processing in familiar talker identification. In Experiment 1, participants identified talkers in three conditions: (a) a foreign-language speech condition featuring unfamiliar sound patterns and no known words; (b) a nonsense speech condition featuring all the familiar sound patterns of their native language, such as familiar phonemes, prosody, and syllable structure, but no actual words; and (c) a native-language condition with all the familiar components of a language, including words. If familiarity with the sound structure of a language allows listeners to perceive subtle idiosyncratic sound productions, listeners would identify talkers better in conditions with familiar sounds than in conditions featuring an unfamiliar language. If lexical knowledge is required to access phonological processes to compare talker idiosyncrasies, listeners would perform better in conditions with familiar words than in conditions with familiar sounds only.

In Experiment 2, participants again identified speakers in familiar and unfamiliar languages. In both languages, listeners identified speakers in a condition in which no word was ever repeated, and in a condition featuring repeated words. As in Experiment 1, listeners were expected to demonstrate higher levels of accuracy identifying speakers in a familiar language than in an unfamiliar language. Additionally, episodic models of speech perception suggest that the accuracy of talker identification would be highest in

the condition featuring access to repeated words in a known language. This increased accuracy would derive from the ability to compare memories of the specific phonetic realization of spoken words, which include idiosyncratic features related to talker identity. Higher levels of accuracy for stimuli with repeated words would only be demonstrated in a familiar language, as listeners would not have abstract lexical representations of words in an unfamiliar language against which to compare memories of talkers' voices. Together, Experiments 1 and 2 explore how linguistic and voice processing systems may be integrated, and how access to different levels of linguistic information may impact talker identification.

The Development of Talker Identification

Children attend to individual speaker characteristics, and may be more sensitive than adults to variability between speakers. Fetuses respond to the voices of their mothers played on tape differently than they respond to other speakers' voices (Kisilevsky, Hains, Lee, Xie, Huang, Ye, Zhang, & Wang, 2003). Similarly, neonates respond differently to their mothers' voices than the voices of strangers, indicating that some familiar talker identification precedes adult-like knowledge of a language (Kisilevsky et al., 2003). Infants begin mapping voices to people other than their mother between 6–8 months, and can discriminate between unfamiliar speakers in a native language between 7–8 months (Houston & Jusczyk, 2003).

At eight months, infants appear incapable of generalizing familiar words to new talkers, indicating extreme sensitivity to differences between speakers (Houston & Jusczyk, 2003). Children's perception is refined as they acquire a native language, and

they generalize familiar words to different talkers at 10.5 months and to unfamiliar dialects at one year (Houston & Juscyzk, 2003; Schmale, Cristia, Seidl, & Johnson, 2010). At fourteen months, hearing the same words in different voices facilitates learning new words, perhaps by drawing attention to the most essential components of a word (Rost & McMurray, 2010). Children between three and five years old struggle to disregard unfamiliar talkers' gender in word classification tasks, and are stronger at distinguishing between familiar speakers, speakers with dissimilar accents, and speakers their own age (Spence, Rollins, & Jerger, 2002; Creel & Jimenez, 2012; Mann, Diamond, & Carey, 1979). By age 5, children successfully match two speakers to faces after three trials with feedback, suggesting the development of fast-mapping may support children's talker identification (Moher, Feigenson, & Halberda, 2010). The negative impact of talker variability on children's talker identification decreases over time, with children identifying speakers at adult levels by age 10 (Ryalls & Pisoni, 1997; Mann, Diamond, & Carey, 1979; Levi & Schwartz, 2013). While infants are substantially more sensitive to talker characteristics than adults, adults retain sufficient sensitivity to talker variation to accurately distinguish between different talkers in their native language (Houston & Jusczyk, 2003; Perrachione, Chiao, & Wong, 2010).

Neuroimaging Studies of Talker Identification

Dedicated neural areas in the superior temporal sulcus (STS) are implicated in the processing of voice, speech, and linguistic information (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). More STS activation is observed in response to human vocalizations than either animal vocalizations or environmental sounds, suggesting mechanisms for human-

voice specific processing (Fecteau, Armony, Joanette, & Belin, 2004). By comparing responses to speech and non-linguistic vocalizations, the processing of vocal sounds was located in right planum temporale and mid anterior STS areas (Capilla, Belin, & Gross, 2013). Familiar and non-familiar speaker identification is associated with posterior STS activation, which is implicated in both sound and non-linguistic voice processing (von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003).

The existence of a language-familiarity effect suggests neural networks for speech and voice processing may be integrated. However, the nature of neuroimaging research is often to isolate functionally distinct regions (Perrachione, Pierrehumbert, & Wong, 2009). Many studies instruct participants to attend to just one feature of stimuli at a time, such as speaker identity or word meaning, and compare levels of activation in order to isolate functionally distinct cortical areas. As a result, studies often highlight distinct areas associated with discrete functions, and may inadvertently obscure overlapping and interconnected regions associated with multiple functions. For example, investigations into the neural mechanisms supporting voice recognition found that substrates involved in recognizing speakers are distinct from areas that process general acoustic information (Belin, Fecteau, & Bedard, 2004). The neural areas implicated in processing linguistic content and speaker information, while closely related, are also associated with different levels of activation (von Kriegstein et al., 2003). Specifically, anterior STS activation is implicated more in talker identification tasks than in the processing of linguistic content, suggesting that the nature of the listening tasks modulates STS activation (von Kreigstein et al., 2003). Similarly, electroencephalographic (EEG) responses to speech sounds when

listeners are instructed to attend to talker identity are slightly different than those observed when listeners are instructed to attend linguistic content (Bonte, Valente, & Formisano, 2009). More right anterior STS activation is measured in response to syllables spoken by different speakers than the same speaker, suggesting this area is involved specifically in the representation of speaker's voices (Belin & Zatorre, 2003). In addition to anterior STS activation, middle and superior temporal and superior parietal regions are associated with the processing required to resolve the phonetic ambiguity introduced by talker variation (Wong, Nusbaum, & Small, 2006). Subsequent neuroimaging identified anatomically-separate areas sensitive to either voice-acoustic changes or changes in voice identity (Andics, McQueen, Petersson, Gal, Rudas, & Vidnyansky, 2010). Whereas right-lateralized auditory areas of the temporal lobe are associated with acoustic processing, left-lateralized temporal areas are thought to be involved in voice identity processing (Andics et al., 2010). Equal levels of right STS activation are observed in response to speech and non-speech vocalizations; however, more left anterior STS activation is measured in response to vocalizations containing linguistic information (Belin, Zatorre, & Ahad, 2002). Additional research suggests the neural processes associated with retrieval of voice identity are distinct (Relander & Rama, 2009) but parallel (Knösche, Lattner, Maess, Schauer, & Friederici, 2002) to those recruited for the processing of linguistic information. By design, these studies do not identify activation common to the processing of both linguistic and speaker information. Although research isolates areas dedicated to voice and linguistic processing, they do not show the role, if any, of simultaneous input from both processes (*cf.* Perrachione, Pierrehumbert, & Wong, 2009).

Using similar methods, researchers isolated areas associated with the processing of familiar voices compared to unfamiliar voices. The electrophysiological responses elicited by a familiar voice are significantly different than those elicited by an unfamiliar voice, potentially allowing for easier perception of familiar voices compared to unfamiliar voices (Beauchemin, de Beaumont, Vannasing, Turcotte, Arcand, Belin, & Lassonde, 2006). Exposure to individual speakers creates dynamic representations of specific voices in long-term memory (Andics et al., 2010). Using event-related potentials (ERPs), researchers located the processing of speaker identity in the auditory cortex within 250ms after the onset of a speech stimulus (Schweinberger, Walther, Zäske, & Kovács, 2011). Recognition of a familiar voice is associated with neural activity in the temporal area, specialized for voices, and the retrosplenial cortex in the posterior cingulate cortex, which is implicated in autobiographical memory and emotional salience (Shah, Marshall, Zafiris, Schwab, Zilles, Markowitsch, & Fink, 2001). A study using short-term fMRI adaptation showed that both lexical information, which relates to the meaning conveyed through speech, and indexical information, which relates to talker identity, are associated with left posterior middle gyrus activation (Chandrasekaran, Chan, & Wong, 2011). Imaging studies of episodic and semantic autobiographical memory, which encompasses an array of processes associated with re-experiencing the past, also found evidence of the involvement of frontal, temporal, posterior, and subcortical structures (Svoboda, McKinnon, & Levine, 2006).

In individuals with developmental phonagnosia, a deficit in speaker identification, abilities such as emotion recognition, speech intelligibility (Garrido et al., 2009;

Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010) and face recognition are unaffected (Roswandowitz et al., 2014), suggesting speaker identification can be isolated from these other processes. Consistent with studies of the neural correlates of familiar talker identification (Shah et al., 2001), acquired difficulty identifying familiar speakers is commonly attributed to right temporal lesions (Gainotti, 2011). Although voice recognition disorders following traumatic brain injury are typically accompanied by impairments to familiar face recognition, speaker recognition disorders are associated more with right STG involvement, and face recognition disorders with right fusiform gyrus involvement (Gainotti, 2011). The association of right STG impairment with phonagnosia is consistent with studies linking right STG activation to the processing of voice, speech, and linguistic information (Belin et al., 2000; Belin & Zatorre, 2003; Fecteau, Armony, Joanette, & Belin, 2004).

Behavioral Studies of Talker Identification

Behavioral studies provide evidence of the integration of neural systems for voice and speech processing. A paper aimed at identifying the functional relationship between speech and voice processing found evidence that linguistic proficiency is required for accurate talker identification (Perrachione & Wong, 2007). In this study, listeners identified speakers in a native and foreign language. Following six consecutive daily opportunities to practice identifying five speakers saying the same five sentences, only listeners with some knowledge of the unfamiliar language achieved the accuracy reached in their native language. Further evidence that language background affects talker identification was found by comparing the ability of monolingual and bilingual listeners

to identify speakers in their first and second language (Bregman & Creel, 2014). Researchers measured the accuracy of talker identification as a function of listener age of language acquisition and familiarity with the language spoken. Bilingual listeners learned to recognize talkers more rapidly in the language they first acquired than in their second language. Additionally, the age of second language acquisition correlated with the learning rate for speakers of the listeners' second language, with earlier acquisition resulting in faster learning. Similar evidence was found in a dichotic listening study investigating the respective contributions of the left and right hemisphere to talker identification (Perrachione, Pierrehumbert, & Wong, 2009). The study compared how dichotic performance predicts binaural performance on talker identification tasks in a native and foreign language. Researchers found right-ear/left hemisphere performance was a better predictor of binaural performance than left-ear/right hemisphere performance in the native language condition, suggesting that the language-dominant hemisphere plays an important role in familiar talker identification in a native language. This relative right-ear advantage when listening to voices speaking a familiar language indicates that the integration of information from language processing is crucial to the language-familiarity effect. This is further supported by evidence that language impairments may result in difficulty identifying voices in native language. Dyslexia, a disorder characterized by difficulty reading, is associated with impaired native-language talker identification but unimpaired identification in a foreign language (Perrachione, Del Tufo, & Gabrieli, 2011). Together, these studies suggest that the integration of the neural networks associated with voice and speech perception is crucial to familiar talker

identification.

As suggested from neuroimaging studies, the processes involved in encoding phonetic representations are closely related to the processes involved in encoding information about a talker's voice (Mullennix & Pisoni, 1990). To comprehend utterances by different talkers, listeners must contend with variability such as overlap in phonological categories across talkers and sounds (Hillenbrand, Getty, Clark, & Wheeler, 1995). By increasing the variability of the non-attended features of a speech signal, researchers increased listeners' reaction time to the other attended features of speech (Mullennix & Pisoni, 1990). Evidence that more time is required to process speech produced by multiple talkers than speech produced by single talkers further indicates that linguistic and talker processing are closely tied (Mullennix & Pisoni, 1990). Additionally, individuals listening to speech embedded in noise more accurately recognize words if they are produced by a familiar talker (Nygaard, Sommers, & Pisoni, 1994), even if the sentences themselves are novel (Nygaard & Pisoni, 1998).

The impact of listener’s beliefs about familiar speakers on speech perception further suggests that linguistic processing and talker identification are intertwined. Listeners perceive acoustically identical vowels as different when they believe they are listening to different speakers, indicating that perceived speaker identity is an important component of vowel normalization (Johnson, 1990). Similarly, listeners expect to hear voice-onset times consistent with a specific talkers' voice-onset time, and use knowledge of talker-specific voice-onset times to facilitate speech processing (Allen & Miller, 2004). Encoding talker-specific acoustic and semantic information allows listeners to

associate familiar talkers with specific referents, thereby facilitating language processing (Creel & Tumlin, 2011). Experience with specific talkers also influences how meaning is derived from speech via the interpretation of structurally-ambiguous sentences (Kamide, 2012). Consistent with episodic theories of speech perception, voice-specific attributes of individual voices may be retained in long-term memory, and have been shown to assist in word recall (Goh, 2005). Listeners expecting a single talker or receiving no instructions showed poorer performance when hearing two different talkers than listeners expecting to hear two different talkers, suggesting listeners actively accommodated talker variability (Magnuson & Nusbaum, 2007). When asked to identify previously spoken words, listener recognition was negatively impacted by changes to talker and rate of speech, both of which impact the phonetic content of an utterance. That changes to amplitude, which does not have an impact on phonetics, did not impact listener recognition suggests that phonetic content plays an important role in word recall (Bradlow, Nygaard, & Pisoni, 1999).

The phonetic variability that impacts speech perception and encoding facilitates talker identification. For listeners to be aware of phonetic variability between talkers, they must be familiar with the sounds of a language. Language familiarity impacts listener's perception of sounds as within a phonetic category or belonging to two different categories (Best, 1994). Without knowledge of a language, it is difficult to determine if specific variations indicate meaningful linguistic differences or acceptable talker variation. In multiple experiments, listeners identify voices better in familiar languages than unfamiliar languages (Thompson, 1987; Goggin, Thompson, Strube, & Simental,

1991). Some researchers posit that this difference is accounted for by limited exposure to the sound structure of a language, rather than limited comprehension of the language. For example, Fleming, Giordano, Caldara, and Belin (2014) asked Mandarin-speaking and English-speaking adults to rate the speaker dissimilarity of pairs of time-reversed sentences in English and Mandarin. Both English-speaking and Mandarin-speaking listeners rated different speakers as more dissimilar when presented with time-reversed recordings in their native language than in an unfamiliar language. As sentences in this study were time-reversed and contained no words, Fleming et al. propose this effect indicates familiarity with a language's "phonology" underlies the language-familiarity effect. However, this study only measured listener judgments of speaker similarity and dissimilarity. As listeners were not asked to identify speakers, the study's applicability to understanding the psychological basis of the language-familiarity effect is limited. In contrast, other studies suggest that talker identification requires linguistic proficiency (Perrachione & Wong, 2007), and that increased linguistic experience in a foreign language results in an increased language-familiarity effect (Creel & Bregman, 2014). These studies support the hypothesis that linguistic knowledge, and not exposure, is responsible for the language-familiarity effect. Additional research is required to determine the specific contributions of language knowledge to talker identification.

The Experiments

This project explored how familiarity with a language facilitates talker identification. Experiment 1 examined participants' ability to identify voices in conditions of varying linguistic familiarity. Native English speakers identified voices speaking in

English, voices speaking in phonologically balanced non-word sentences, and voices speaking in an unfamiliar language (Mandarin Chinese). In order to be as similar to English as possible, the non-word English sentences contained the sounds, syllables, syllable structure, and prosody of English, but did not contain English words. While it was expected that participants would more accurately identify voices in their native language than in an unfamiliar language, there were several possible outcomes of the non-word sentences condition. Listener familiarity with the sounds in the non-word sentences could result in talker identification accuracy similar to the English sentence condition. This would indicate that the native-language benefit derives from familiarity with the sound structure, and that familiar words themselves do not enhance talker identification accuracy. Alternately, should access to familiar words facilitate talker identification, the listeners' accuracy at talker identification in the non-word sentences would be similar to the unfamiliar language condition. Finally, listener accuracy in the non-word condition could fall between the accuracy achieved in the other two conditions. While this outcome would indicate that access to all the familiar components of a language contribute to an individual's ability to identify talkers in an additive relationship, it would also suggest that access to words is an especially important component of the language-familiarity effect.

Experiment 2 investigated the role of memory for words in identifying familiar speakers by exploring the contributions of word repetition to talker identification. As in Experiment 1, participants identified talkers speaking sentences in English and in an unfamiliar language. The speakers of both English and the unfamiliar language recorded

a set of sentences in which no word was ever repeated. Participants identified speakers saying repeated sentences in English, repeated sentences in an unfamiliar language, unrepeated sentences in English, and unrepeated sentences in an unfamiliar language. In the unrepeated sentences condition, participants could not use episodic memory of a speaker's production of specific words to facilitate talker identification, and could not compare productions of specific words across speakers. However, listeners could use different words to access familiar phonology in the English unrepeated sentences condition. Together, Experiments 1 and 2 tested talker identification in conditions that provide different amounts of linguistic information to explore the specific linguistic knowledge that supports talker identification.

EXPERIMENT 1

Participants

A group of Native English-speaking adults (N = 24; 7 male) with a self-reported history free of speech, language, hearing, and learning impairments participated in this experiment. Participants were unfamiliar with both the voices recorded for use as stimuli in the experiment and Mandarin Chinese, and performed above chance (i.e., 20%) in all conditions. An additional six participants were recruited; however, they were excluded following failure to demonstrate a language-familiarity effect. All participants gave written, informed consent, and were paid for their participation. None of the participants in Experiment 1 participated in Experiment 2. The subject pool was recruited through advertisements on Boston University job boards.

Stimuli

Participants learned the voices of five speakers of English sentences, five speakers of English non-word sentences, and five speakers of Mandarin Chinese sentences. Ten female speakers of American English (age 20–29, M = 23 years) with similar regional accents recorded 60 English sentences from six lists of the Harvard sentences (IEEE, 1969) and 60 English non-word sentences (See Appendix B). The English non-word sentences are phonologically valid and phonotactically matched to the English sentences. They were developed by transcribing the six lists of ten Harvard sentences phonemically with broad phonemic transcription, and then rearranging the phonemes within each list to form ten new nonsense sentences. These sentences contain zero English words, but do

not differ from the English sentences in length, number of syllables, phoneme positional probability, or biphone positional probability. The speakers who recorded these English and non-word sentences all had training in phonetics, were extensively familiarized with the target pronunciation of the non-word sentences, and were recorded multiple times to ensure fluency. Additionally, ten female native speakers of Mandarin Chinese (age 18–36, M = 26 years) with regionally homogeneous accents recorded a set of phonetically-balanced Mandarin sentences from the "Mandarin Speech Perception Tests" (Fu et al., 2011). No talkers were recorded in both languages or participated in the experiment. All recordings were made at 44.1 kHz in a sound-attenuated chamber using a Shure SM58 vocal microphone, Behringer MIC2200 microphone preamplifier, and FCA1616 USB sound card, and normalized to 65 dB SPL RMS amplitude. Recording durations were 1.83 ± 0.25s in English, 1.97 ± 0.19s in non-word English, and 1.58 ± 0.21s in Mandarin Chinese. Recordings, acoustic processing, and acoustic analysis were conducted in Praat.

English Sentences	Mandarin sentences	Non-word sentences
The boy was there when the sun rose.	今天的阳光真好 jīn tiān de yáng guāng zhēn hǎo *It's a nice sunny day.*	dɔI əz səv strɑl həp pəvsmɑIrs
A rod is used to catch pink salmon.	晚上一块去跳舞 wǎn shàng yī kuài qù tiào wǔ *Let's go dancing together tonight*	rɑft zis tɝl zəb jɛtʃ kudʒ keib
The source of the huge river is the clear spring.	对面有两所高中 duì miàn yǒu liǎng suǒ gāo zhōng *There are two high schools across the street.*	wɛk ɑt oki ʊʃn brɑlt gæt ð^n

Table 1: Sample stimuli. This table displays three examples of sentences from each language condition. English translation of the Mandarin sentences is also provided.

Design and Procedures

Participants were trained and tested in talker identification in three conditions. Speakers in each condition were grouped into two sets of five speakers to maximize perceived voice distinctiveness based on extensive piloting. Each voice was associated with a number and a semantically and visually distinctive clip art avatar. Different avatars were used in each condition to minimize confusion. The order of conditions was counterbalanced across experiments.

The experiment was designed in PsychoPy (Peirce, 2007). Each participant was seated in front of a computer screen to view stimuli with PsychoPy (v1.8.0) at a comfortable listening level via Sennheiser HD 380 Pro circumaural headphones in a sound attenuated booth. Participants responded by selecting numbers corresponding to speakers on a keypad. The experiment was self-paced. Participants completed training and testing in one language condition prior to beginning the next language condition. Each condition featured two phases: Familiarization and Test (See Figure 1).

Familiarization

At the start of the Familiarization phase, the following instructions appeared on the screen:

> "In this part of the experiment you will meet five different people who are speaking English. Your job is to learn to identify who is who. It doesn't matter what they are saying, because the only thing you need to pay attention to is which person is talking. If you understand these directions, press the space bar to continue. Now you will be introduced to the five speakers. Let's practice. Press

the appropriate button to indicate which person you hear talking. The computer will help you if you make a mistake. If you understand these directions, press zero to continue."

On the next screen, participants passively heard one speaker at a time say the same sentence, while the corresponding avatar and number appeared on the screen. Next, all speakers' avatars and numbers appeared on the screen while one of the speakers repeated the sentence participants just heard from all speakers. Participants identified who they believed was speaking by pressing the number that corresponded to the speaker number on a keypad. After each trial, participants received feedback on their choice. The screen displayed the correct avatar and number, and either "Correct" or "Incorrect. The correct number was X." This was repeated until each talker had said the sentence two times. Training was then repeated for the next sentence, until listeners had practiced identifying talkers twice for all five different sentences, for a total of 50 exposure trials and 50 quiz trials with feedback. This procedure was the same for the American English, Mandarin Chinese, and English non-word sentence conditions.

Test

After participants were familiarized with each of the five talkers, the Test phase began. The following instructions appeared on the screen:

"Now there will be a test. Press the appropriate button to indicate which person you hear talking. The computer will not tell you the right answer this time. If you understand these directions, press the space bar to continue."

Trials in the test phase were the same as the Familiarization phase; however,

feedback was not provided. Participants identified the five talkers saying each of five sentences two times, for a total of fifty sentences.

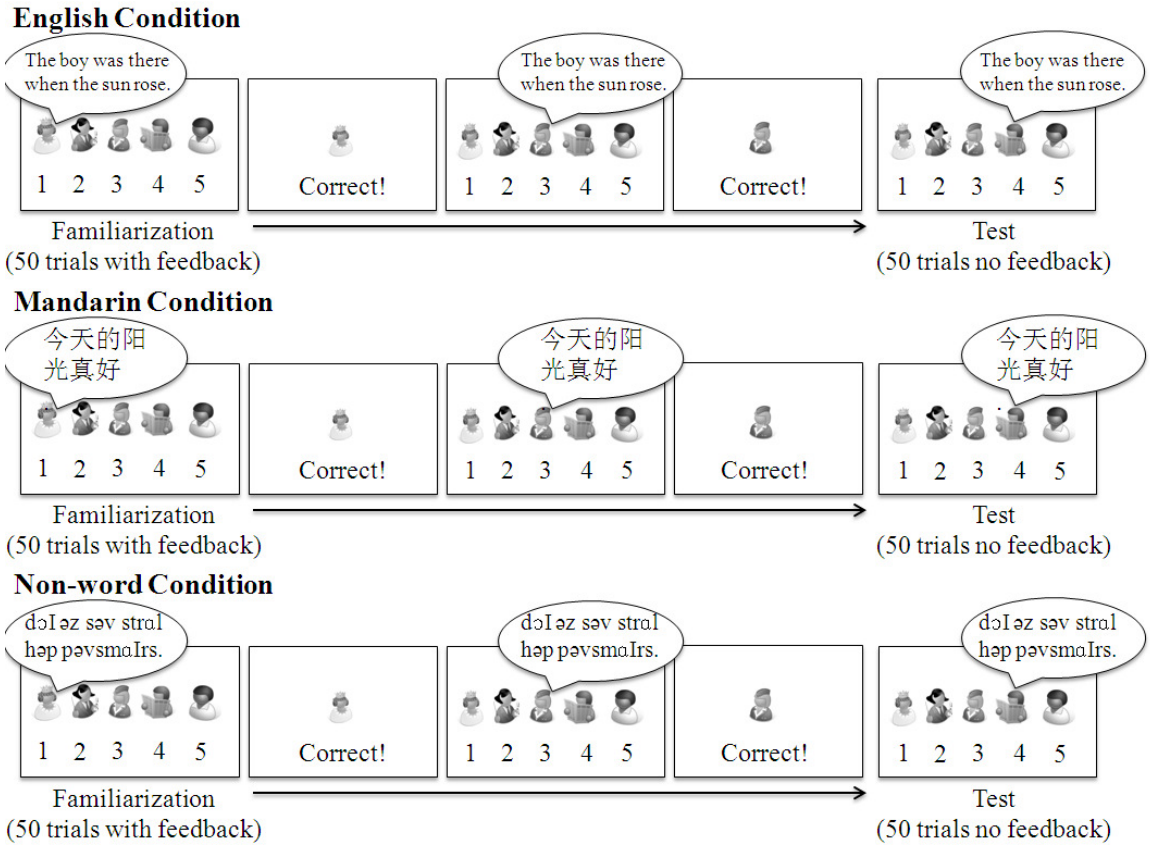

Figure 1: A schematic diagram of the phases of Experiment 1 in each condition.

Results

As this experiment examines the influence of lexical and phonetic information in the language-familiarity effect, participants not demonstrating a language-familiarity effect (i.e., percent accuracy was higher in Mandarin Chinese than English) were excluded. Participants' accuracy on the test phases of the English, non-word, and Mandarin Chinese conditions were analyzed in R using generalized linear mixed effects

models for binomial data with Condition as the fixed factor. A maximal random effects structure included random intercepts and slopes by participants, with random intercepts by talker (Barr, Levy, Scheepers, & Tilly, 2013). Due to the enhanced voice recognition abilities of participants in their native language**,** participants most accurately identified talkers in the English condition (71.3% ± 15%), and least accurately in the Mandarin condition (43.5 ± 13%). Accuracy in the Nonword condition (64.3 ± 14%) surpassed the Mandarin condition but fell short of English accuracy. The difference in accuracy between each of these three conditions was significant (English/Mandarin: $z = 4.79$, $p < 2\times10^{-6}$, Cohen's $d = 1.98$; English/Nonwords: $z = 2.71$, $p < 0.007$, $d = 0.49$; and, Nonwords/Mandarin: $z = 3.75$, $p < 0.0002$, $d = 1.57$). These results are depicted in Figure 2.

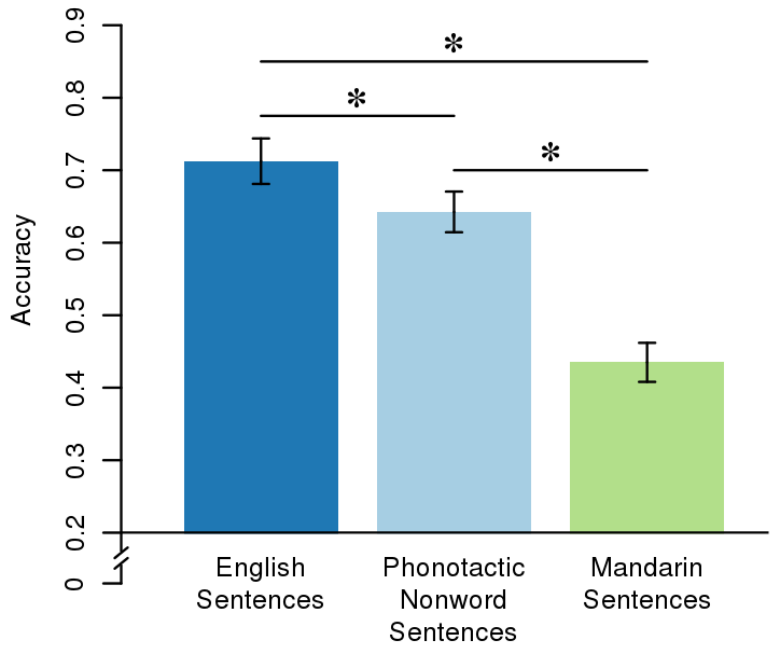

Figure 2: Average listener accuracy at talker identification in each condition. Listener accuracy improved with increased access to familiar linguistic information. Access to familiar words provided an advantage over stimuli that are as familiar as possible without containing real words. Error bars indicate standard error of the mean.

Discussion

The results indicate that talker identification is supported by access to both familiar sounds and familiar words. As anticipated, listeners identified talkers least accurately when they heard a language featuring an unfamiliar sound structure and no familiar words. Accuracy improved over a foreign language when listeners heard a meaningless language featuring familiar sounds, syllables, prosody, and sound structure. This finding suggests that, even in the absence of familiar words, other aspects of a language contribute to the language-familiarity effect.

However, the advantage of access to familiar words persists even compared to stimuli that are as familiar as possible without containing real words. This finding reveals the importance of mental representations of words to accurate talker identification. Listeners appear to benefit from the ability to compare episodic representations of words to heard words to recognize idiosyncratic talker characteristics. This benefit is not observed when listeners hear words in an unfamiliar language. Listeners have no prior representations of words in an unknown language in their mental lexicon, and have stored neither the abstract concept nor specific features of its spoken production. The increased accuracy observed when familiar words are available indicates that lexical access derived from meaningful speech contributes to the language-familiarity effect. The effect supports an integrated model of voice and speech processing.

EXPERIMENT 2

Participants

A new group of native English-speaking adults (N = 16; 4 male) participated in this experiment. As in Experiment 1, all participants met a series of inclusion criteria, including: demonstration of a language-familiarity effect (i.e., better performance in English than Mandarin conditions), unfamiliarity with the voices in the experiment, lack of prior experience with Mandarin Chinese, performance above chance (i.e., 20%) in all conditions, and a self-reported a history free of speech, language, hearing, and learning impairments. Six participants were excluded following failure to demonstrate a language-familiarity effect. None of the participants in Experiment 2 had previously participated in Experiment 1. All participants provided written informed consent, and received monetary compensation for their participation.

Stimuli

Two sets of stimuli were created for Experiment 2. The first set consisted of 100 five to eight syllable phonotactically-balanced sentences (M = 7) in which no word is repeated within sentences or across the entire set (See Appendix A). Ten female native speakers of American English with similar regional accents (age 20–29; M = 23 years) were recorded saying all 100 of the sentences (average duration: 1.66 ± 0.18s). Additionally, ten female native speakers of Mandarin Chinese were recorded saying 100 phonotactically-balanced seven syllable sentences in which characters were very rarely repeated (average duration: 1.58 ± 0.09s) (Fu, Zhu, & Wang, 2011). The speakers of

Mandarin Chinese (age 18–36; M = 26) also had regionally similar accents. No talkers were recorded in both language conditions or participated in the experiment. All recordings were made at 44.1 kHz in a sound-attenuated chamber using a Shure SM58 vocal microphone, Behringer MIC2200 microphone preamplifier, and FCA1616 USB sound card, and normalized to 65 dB SPL RMS amplitude. Each voice was associated with a semantically and visually distinctive clipart avatar of a person.

Design and Procedures

Participants learned to recognize voices in four conditions in a 2x2 design. The first condition assessed talker identification of speakers of American English saying sentences in which no word was ever repeated. The second condition assessed talker identification of speakers of American English in which sentences were repeated across speakers. The remaining two conditions were the same as the first two; however, the talkers spoke in Mandarin Chinese.

Voices were counterbalanced across participants and experiments. The order of conditions was also counterbalanced across participants. Each participant was seated in front of a computer screen for this self-paced experiment. Stimuli were presented with PsychoPy (v1.8.0) at a comfortable listing level using Sennheiser HD 380 Pro circumaural headphones in a sound attenuated chamber.

Training and testing were completed in one language condition before proceeding to the next language condition. Three were two phases: Familiarization and Test (See Figure 3).

Familiarization

At the start of the Familiarization phase, the following instructions appeared on the screen:

> "In this part of the experiment you will meet five different people who are speaking English. Your job is to learn to identify who is who. It doesn't matter what they are saying, because the only thing you need to pay attention to is which person you hear talking. The computer will help you if you make a mistake. If you understand these directions, press zero to continue."

On the next screen, participants heard one speaker say a sentence while all the speakers' avatars and numbers appeared on the screen. Participants identified who they believed was speaking by pressing a number on a keypad that corresponded to the speaker number. Participants received feedback on their choices, and were shown the avatar and number of the correct speaker if they responded incorrectly. The screen displayed the correct avatar and number, and either "Correct" or "Incorrect. The correct number was X." This cycle was repeated for a total of 60 trials. In the repeated sentences condition, all five voices said the same 12 sentences. In the unrepeated sentences condition, five voices each said 12 unique sentences, with no sentences or words ever repeated. Sentences and speakers occurred in a computer-randomized order.

This procedure was the same for the American English and Mandarin Chinese conditions.

Test

The Test phase closely resembled the Familiarization phase, but featured no

feedback in testing. The following instructions appeared on screen:

> "Now there will be a test. Press the appropriate button to indicate which person you hear talking. The computer will not tell you the right answer this time. If you understand these directions, press zero to continue."

There were 25 total trials in the test phase. In the unrepeated condition, each of the five speakers said five different sentences, for a total of twenty-five sentences in which no sentences or words were repeated within or across talker. None of the test sentences or words were previously heard in the Familiarization phase. In the repeated sentences condition, the five sentences were randomly selected from among the twelve sentences said in the Familiarization phase. The participants heard each of the same five sentences read by all five speakers for a total of twenty-five trials.

This experiment was repeated with trials in Mandarin Chinese in the same format.

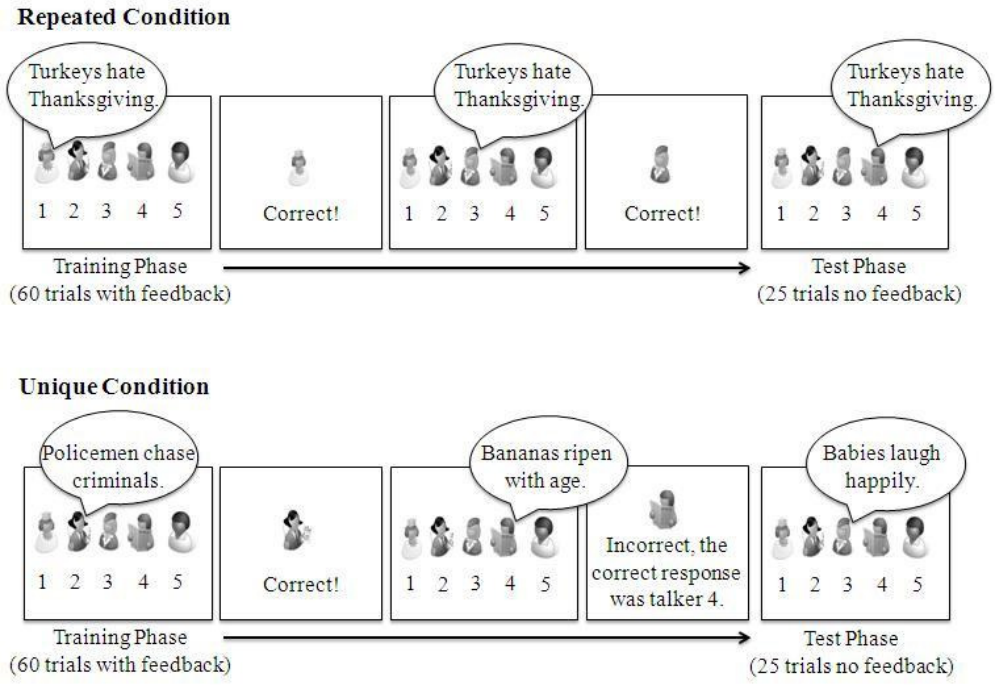

Figure 3: A schematic diagram of the phases of Experiment 2 in each condition.

Results

Word repetition and talker identification

As in Experiment 1, participant accuracy scores were calculated in R using a generalized linear mixed effects model for binomial data. Fixed factors included Language and Condition, and a maximal random effects structure included random intercepts by talker, and random intercepts and slopes by participant (Barr et al., 2013). There was a significant effect of Repeated versus Unrepeated condition (z = 4.22, p = 2.5×10^{-5}), with participants demonstrating greater accuracy when presented with talkers saying the same sentences than talkers saying unique, unrepeated sentences. Additionally, there was an interaction between language and condition (z = 3.26, p = 0.0011), revealing a greater effect of condition in English than in Mandarin. Due to this interaction, the generalized linear mixed effects model was repeated using data from each language separately. This analysis revealed that while listeners identified speakers more accurately in the repeated sentences condition in English (z = 4.23, p = 2.4 x 10^{-5}), the Mandarin talkers were not identified more accurately in the repeated sentences condition (z = 0.23, p = 0.82).

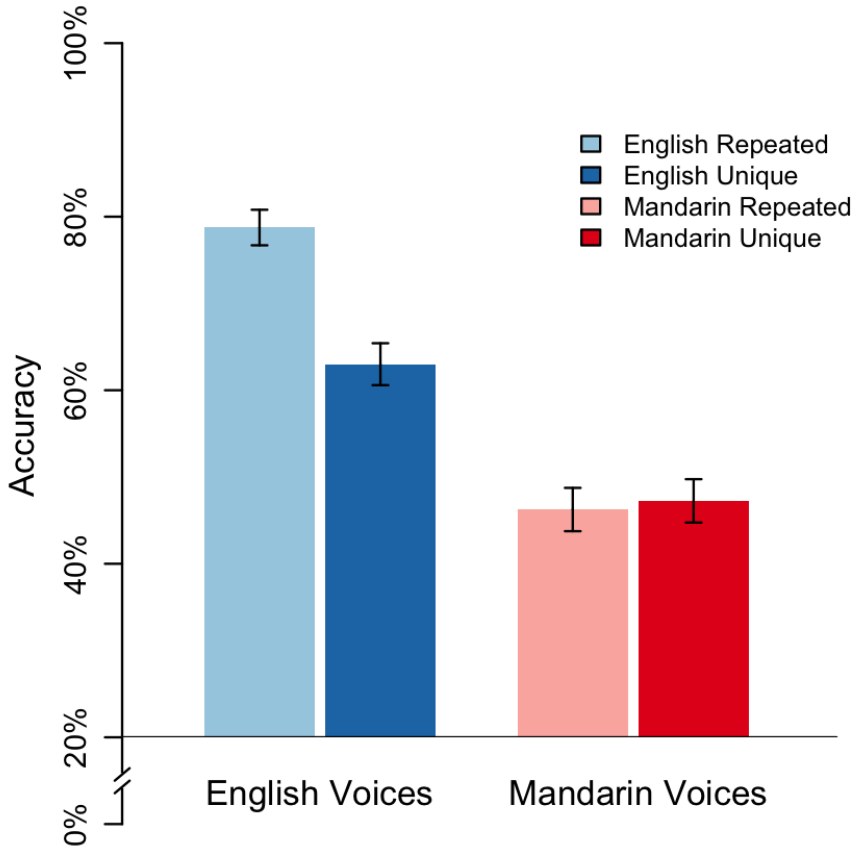

Figure 4: Overall participant accuracy at talker identification in each condition. Increased accuracy was achieved in English compared to Mandarin conditions. Greater accuracy in the repeated words condition compared to unique words was demonstrated in the English condition only. Differences in the levels of accuracy demonstrate the benefit of hearing repeated words for talker identification in a known language. Error bars represent standard error of the mean.

Word repetition and the language-familiarity effect

Listeners always identified voices better in English than in Mandarin, as evidenced by an overall effect of language for each condition (Repeated: $z = 6.28$, $p < 3.5 \times 10^{-10}$; Unique: $z = 2.20$, $p = 0.028$), as well as in the full model ($z = 2.48$, $p = 0.013$). The magnitude of each participant's language-familiarity effect in each condition was calculated as the difference between their performance identifying English and Mandarin voices. The effect size of the language-familiarity effect was significantly larger in the

Repeated condition (d = 2.61) than in the Unrepeated condition (d = 1.12). The magnitude of the language-familiarity effect was greater when talkers said the same words than when no words were repeated (t_{15} = 3.12, p = 0.007).

Discussion

The results of Experiment 2 confirm that memory for linguistic content supports talker identification. Listeners appear to use episodic memories of spoken words to compare and recognize distinguishing features of different talkers. Repeated access to the same words supports talker identification, but only if that language is understood. The absence of a benefit from hearing repeated words when talkers spoke an unfamiliar language indicates that episodic memories support talker identification only when listeners can access meaningful words. In order to benefit from the word repetitions, listeners must have access stored episodic representations of meaningful words, which are then compared to the word they are currently hearing.

GENERAL DISCUSSION

Models of speech and voice processing

The results of Experiments 1 and 2 are consistent with episodic models of lexical access. As discussed, episodic models argue that representations of words include aggregated information from every occasion a word is spoken (Goldinger, 1998). In both Experiments 1 and 2, listeners achieved the highest levels of accuracy when identifying talkers speaking familiar words. In Experiment 2, in particular, listeners most accurately identified speakers who said the same words multiple times. While the features of a familiar language, such as prosody and syllable structure, assist in talker identification, access to familiar words is essential to achieve maximum levels of accuracy. Listeners appear to compare the memories of a specific spoken word to other instances of the same spoken word in order to better identify the speaker. As predicted by episodic models, the benefit of access to repeated words is only observed in a known language for which listeners could access meaningful memories of words. Episodic experiences with speech were likely not encoded as deeply when they did not involve known lexical items, resulting in no benefit of repetition in the foreign language condition.

The results of this project contradict the proposal by Fleming et al. (2014) that exposure to the sound structure of a language is sufficient to account for the language-familiarity effect. Rather, the results build on earlier findings that the language-dominant hemisphere plays a role in talker identification in a known language (Perrachione, Pierrehumbert, & Wong, 2009). Specifically, they support evidence that linguistic background (Bregman & Creel, 2014) and proficiency (Perrachione & Wong, 2007)

directly impact the accuracy of talker identification. The importance of lexical access to the talker-familiarity effect also helps to account for the low native-language talker identification accuracy demonstrated by individuals with dyslexia (Perrachione, Del Tufo, & Gabrieli, 2011). Difficulty encoding, storing, or accessing accurate episodic representations of words could potentially account for difficulties both reading words and identifying talkers.

The role of words in the language-familiarity effect is suggestive of an integrated system for voice and speech processing. Evidence of this intertwined processing must be reconciled with neural evidence of specialized mechanisms for the processing of speech and voice information. The methodology of neuroimaging studies, which often identify differences in the processing of linguistic and speaker information, may have inadvertently obscured the integration of voice and speech processing (Perrachione, Pierrehumbert, & Wong, 2009). By comparing activation associated with different tasks, studies could not locate activation common to both speech and voice processing (Perrachione, Pierrehumbert, & Wong, 2009). Differences in right anterior STS activation for different tasks, including speech and voice processing (Bonte, Valente, & Formisano, 2009) and vocal acoustic and vocal identity processing (Andics et al., 2010), do not preclude the additional presence of integrated processing for speech and voice information (Perrachione, Pierrehumbert, & Wong, 2009). Evidence that lexical information impacts the processing of talker identification is consistent with recent evidence of left posterior middle gyrus activation for the processing of both lexical and indexical information (Chandrasekaran, Chan, & Wong, 2011).

Development of talker identification

Episodic lexical access has implications for the development of talker-identification abilities. By accumulating episodic memories of words, infants may learn to distinguish between word-level differences that carry linguistic meaning and talker-level differences that are not linguistically meaningful (Rost & McMurray, 2010). Episodic theories suggest that infants struggle to generalize familiar words to new talkers (Houston & Jusczyk, 2003) because they have yet to accrue sufficient episodic memories of spoken words. The importance of episodic representations of words to talker identification also helps explain why the emergence of fast-mapping coincides with an increase in the ability to identify speakers (Moher, Feigenson, & Halberda, 2010). As suggested by Perrachione & Wong (2007), increased linguistic knowledge in the form of more memories of words may allow children to become more proficient at both recognizing familiar words and recognizing talkers by age ten (Levi & Schwartz, 2013).

Clinical implications

The integration of voice and speech processing carries clinical implications for perceptual voice and speech evaluations. Listeners' perceptions of a speaker's voice may vary due to both the content of an utterance and the speaker's vocal quality. Previous experiences with speakers and specific spoken words may influence listeners' assessments of speakers' voices. Given the importance of words to the language-familiarity effect, listeners' perceptions of vocal quality when presented with a sustained /ɑ/ or an utterance spoken in a different language may be less affected by the language-familiarity effect than when presented with a familiar or a repeated utterance.

Additionally, evidence that listeners quickly learn and accommodate the idiosyncratic indexical features of speakers' voices suggests that objective measures of voice and speech may be important in the evaluation and treatment of disorders impacting intelligibility.

Limitations

A number of participants were excluded for failure to demonstrate a language-familiarity effect. The absence of such a well-documented effect for some listeners may limit the generalizability of these results to all scenarios involving talker identification. Correspondingly, they illustrate differences between this and previous studies regarding the design of the stimuli, the effort demanded of listeners, or other so-far unidentified effects of differences in design. Many experiments regarding the language-familiarity effect presented recordings of male speakers (e.g., Thompson, 1987; Goggin et al., 1991; Perrachione & Wong, 2007; Perrachione, Del Tufo, & Gabrieli, 2011), raising the possibility that different factors mediate the identification of female voices. Although studies have found evidence of a language-familiarity effect in languages such as German, Mandarin Chinese, Korean, Spanish, and English (e.g., Goggin et al., 1991; Perrachione & Wong, 2007; Levi, Winters, & Pisoni, 2011; Bregman & Creel, 2014), further study is required to extend the implications of these results to all spoken languages.

CONCLUSION

By testing talker identification in conditions that provide different amounts of linguistic information, this project demonstrated the importance of lexical access to the language-familiarity effect. In Experiment 1, listener accuracy at talker identification improved over a foreign language when listeners heard a non-word language that was as familiar as possible without containing real words; however, access to known words resulted in the highest levels of accuracy overall. In Experiment 2, listeners identified talkers most accurately when given access to repeated known words, further suggesting that episodic memories of spoken words are used to compare and recognize distinguishing features of different talkers. Together, these experiments indicate that adults' high levels of accuracy when identifying talkers in a native language are supported by the integration of voice and speech processing.

APPENDICES

Appendix A: Unique word sentences

granola tastes best in yogurt
striped sweaters sell quickly
telescopes view constellations
babies laugh happily
policemen chase criminals
maps show country boundaries
puppies bark at passing cars
polka dots decorate fabric
students whisper secrets
perennials bloom all year
textbooks burden backpacks
calculators solve problems
chapstick moisturizes lips
handy rulers draw straight lines
coined money crowds my wallet
mineral ice relieves joint pain
studying improves exam scores
reusable bottles save trees
noisy alarms wake roommates
kangaroos jump along
treetops sway backwards and forwards
wind pushes against heavy doors
most flowers grow slowly
unwelcome weeds invade lawns
special coffee mugs are good gifts
thumb tacks support posters
bosses manage employees
plugs supply electricity
city sidewalks dirty shoes
insulation stops heat loss
vacuums clean messy rugs
mops wipe up grimy floors
cellphones beep from text messages
planes fly over oceans
newspapers publish articles

license plates register cars
professors teach lectures
late assignments don't count
stressful exams consume time
paint can revitalize houses
fuzzy slippers warm feet
smelly socks need cleaning
locks protect valuables
fans circulate stagnant air
magnets stick on fridges
coupons lower prices
paintings ornament walls
Japanese lanterns lit dark rooms
microwaves cook meals fast
spatulas flip burgers
bananas ripen with age
shower pipes spray water
children run around open fields
mason jars store jelly
Splenda replaces sugar
insults merit little response
kids collect rough seashells
fluorescent light casts ugly rays
clouds inch across blue skies
recycling bins gather paper
red wine stains white seat cushions
broken headphones produce static
buckles enhance boring boots
slideshows assist presentations
trolleys roll down busy streets
flu symptoms worsen outside
inaccurate clocks cause lateness
benches fill during lunchtime
firemen forbid blocking exits
church steeples point skyward
plants flourish inside greenhouses
fences keep intruders out
sun shines onto window panes
some women wear lipstick

numbers mark room locations
cones avert through traffic
he prefers ballpoint pens
junk emails crowd inboxes
sometimes rooftops could be flat
lunch bags carry sandwiches
chalk boards contain equations
railings make stairways safe
rabbits munch tasty carrots
mantles display awards
owls hoot when frightened
annoying birds chirp noisily
construction workers holler
tots adorn easter eggs
bold girls proposition boys
baseball has our attention
hungry squirrels devour nuts
glasses magnify small print
overdue books deserve fines
long sleeve shirts prevent bug bites
ground hogs destroy gardens
no one likes filthy counters
turkeys hate Thanksgiving
washing machines spin rapidly
mother gives excellent hugs

Appendix B: Phonologically Balanced Non-word Sentences

dɔI əz səv strɑl həp pəv smɑIrs
rɑft zis tɝl zəb jɛtʃ kudʒ keib
wɛk ɑt oki ʊʃn brɑlt gæt ð^n
kɑd həz Ik bæmn hɛs læn brol
ðəm bud rɛn pjuzed Iŋ wʊfit
pæf læsk fɑk kesi sof wɛz
æmIŋk sprors fɑlu gɛk nIŋ
him mæsə krɑm lIf wən tIz
ivɚ θrət tət set ðɑl frəs əm
tɑndz tə sɛr təm ðəlo tət ivət
ə ʃorn sɛpt ɑImɝd lel
skwon rIl hov st^g ɑImr Iŋ
Ip kiknəz ʃod ku ovIft wop
ɑpɚ dɛft wɛʃ dIn wɑrt hod
hIn lɚ grɑʊ sɑIndə ðens gorg
kɑId uld ʃif hIk ð^n læzi stɛn
bɛvn fænd frənd Ins dɝli tet
ɑrp θɝm sə færns ən wʊks plIn
ðəli ips ez ðɑn θɝm rIn lI
bə ^ndʒ tʃI əzɑrt wi rəd bə wæt ^ns
tʃɑs mɑrp wæst fɑʊs roŋ mep
soldʒ ʃʊg plɑtʃ vɛriŋ æŋ ʃəp
lidʒ ʃet ək sɑm tjudʒ swidʒ bovɚ
frɑs set flod mɛŋ nek tʊp
swɑd prop fət lisəm lim wɛdl
tʃep rɑfd hɔI s^t lɑIdʒIŋ plef
k^v əlɑd ɔIlst θædʒ vɛrdz
rɑfd kudʒ must ɔIkst ævəm lorm
drum letʃ məl mɑIp bɑθ retʃ
mɑʊt səθ swist strɑd wɑp bormz
jɑd ðu ðIls ðɑld læn tes
ðorn hɑŋli θIl frup sIld
mət fInɑn tɑʊz ris skwɝ slok
hIθ vɚn wən pɔI wɑIŋ Iŋt
wis ɑʊk mɑkə lɑrn f^p wəl
k^l intə pleIl pedrɑI sind Inəl

sIlə mɑIp kɑw trok pɜ˞n t^t
əraʊntu ɑðə ɑdʒ wIθɑIt stin
mɛpə vIŋ ðɜ˞m wIrIn slɑIk səz
əðe sæn faʊt ðɑrt boŋ stræn tinkə
tɑIsts hɑrs bɑʃ kræt pæski læŋ
Ik bint fiv ðət lɑʊ hig rætʃ ^nd
dɜ˞st wɑ ɑt ækʊd wɑIIŋ ɑIni
θord ɑt ɑIni æpIld bræz Iski
stæp ðəf flet gɜ˞k əvə˞z
rɑfl sInt klɛnst uz rIn ðə˞
tuf pri rɑIdə won sInt hə
sər pɑk ʃorf vik ɑk hɛs
kɑ ɛnts pom rə Inət ðɑ
ə Iski gɜ˞k wɑrv klɛnst d^t
səmɛnd hæn wɛvd Id rɑIdəm uvd
lɑv sɛk əvNdə ʃIk tekəd hɑr
slɑIf meŋ əlɑ wɑ Itl ðəs wem
dut wæθ wʊdI stə wən
lɑIʃ wʊb di kɛk ^və˞d wə˞p
lɑrs hɑʊdʒ hæp tæt wɑp tə˞ sə
hɑrz Id blor təs Irɑ Iŋ tus
rɑIn tæt wəm eI gɜ˞t fe ætor
dord meb nɑn kli bræb rɑIv əz wə
rɛb kɜ˞d bɑIŋk æst rit menɑn ðət

REFERENCES

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

Andics, A., McQueen, J.M., Petersson, K.M., Gal, V., Rudas, G., & Vidnyansky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage, 52*(4), 1528–1540.

Barr, D., Levy, R., Scheepers, C., & Tilly, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Beauchemin, M, de Beaumont, L.D., Vannasing, P., Turcotte, A., Arcand, C., Belin, P, & Lassonde, M. (2006). Electrophysiological markers of voice familiarity. European *Journal of Neuroscience, 23,* 3081–3086.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135. doi: 10.1016/j.tics.2004.01.008

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right temporal lobe. *Neuroreport, 14* (16), 2105–2109.

Belin, P., Zatorre, R.J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*, 17–26.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 307–312.

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

Bonte, M., Valente, G., & Formisano, E. (2009). Dynamic and task dependent encoding of speech and voice by phase reorganization of cortical oscillations. *Journal of Neuroscience, 29*(6), 1699–1706.

Bradlow, A.R., Nygaard, L.C., & Pisoni, D.B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics, 61*(2), 206–219.

Bregman, M.R. & Creel, S.C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*(1), 85–95.

Capilla, A., Belin, P., & Gross, J. (2013). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. *Cerebral Cortex, 23*(6), 1388–1395.

Chandrasekaran, B., Chan, A.H.D., & Wong, P.C.M. (2011). Neural processing of what and who information in speech. *Journal of Cognitive Neuroscience, 23*(10), 2690–2700.

Creel, S.C., & Jimenez, S.R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology, 113*(4), 487–509. doi: 10.1016/j.jecp.2012.07.007

Creel, S.C. & Tumlin, M.A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*(3), 264–285.

Fecteau, S., Armony, J.L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage, 23,* 840–848.

Fleming, D., Giordano, B.L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 111*(38), 13795–13798.

Fu, Q., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *Journal of the Acoustical Society of America, 129*(6), 267–273.

Gainotti, G. (2011). What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia, 49,* 2273–2282.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., & Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123–131.

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.

Goh, W.D. (2005). Talker variability and recognition memory: Instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1): 40–53.

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

Hailstone, J.C., Crutch, S.J., Vestergaard, M.D., Patterson, R.D., & Warren, J.D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia, 48*(4), 1104–1114.

Hillenbrand, J. M., & Houde, R. A. (2003). A narrow band pattern-matching model of vowel perception. *Journal of the Acoustical Society of America*, *113*(2), 1044–1055. doi:10.1121/1.1513647

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1143–54.

Institute of Electrical and Electronics Engineers. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 225–246.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America, 88,* 642–654.

Kamide, Y. (2012). Learning individual talker's structural preferences. *Cognition, 124*(1), 66–71.

Kisilevsky, B.S., Hains, S.M.J., Lee, K., Xie, X., Huang, H.F., Ye, H.H., Zhang, K., & Wang, Z.P. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14*, 220–224.

Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage, 17*(3), 1493–1503.

Kreiman, J., & Sidtis, D. (2011). Recognizing speaker identity from voice: Theoretical and ethological perspectives and a psychological model. In *Foundations of voice studies: An interdisciplinary approach to voice production and perception* (156–188). Malden, MA: Wiley-Blackwell.

Levi, S.V, & Schwartz, R.G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech Language and Hearing Research, 56*(3), 913–920.

Levi, S.V., Winters, S.J., & Pisoni, D.B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *Journal of the Acoustical Society of America, 130*(6), 4053–4062.

Magnuson, J.S., & Nusbaum, H.C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391–409.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition-parallels with face recognition. *Journal of Experimental Child Psychology, 27*(1), 153–165. doi: 10.1016/0022-0965(79)90067-5

Moher, M., Feigenson, L., & Halberda, J. (2010). A one-to-one bias and fast-mapping support preschoolers' learning about faces and voices. *Cognitive Science, 34,* 719–751.

Mullennix, J.W. & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *47*(4), 379–390.

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or projecting sexual orientation through speech. *Language and Linguistics Compass*, *1,* 416–449.

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

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Perrachione, T.K., Chiao, J.Y., & Wong, P.C.M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition, 114*(1), 42–55.

Perrachione, T.K., Del Tufo, S.N., & Gabrieli, J.D.E. (2011). Human voice recognition depends on language ability. *Science, 333,* 595.

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 and Performance, 35,* 1950–1960.

Perrachione, T.K., & Wong, P.C.M. (2007). Learning to recognize speakers of a nonnative language: Implications for the functional organization of human auditory cortex. *Neuropsychologia, 45*(8), 1899–1910. doi: 10.1016/j.neuropsychologia.2006.11.015

Peirce, J.W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.

Relander, K. & Rama, P. (2009). Separate neural processes for retrieval of voice identity and word content in working memory. *Brain Research*, *1252,* 143–151.

Rost, G.C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early world learning. *Infancy, 15*(6), 608–635. doi: 10.1111/j.1532-7078.2010.00033.x

Roswandowitz, C., Mathias, S.R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology, 6*(19): 2348–53. doi: 10.1016/j.cub.2014.08.048

Ryalls, B.O., & Pisoni, D.B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33*(3), 441–452.

Schmale, R., Cristia, A., Seidl, A., & Johnson, E.K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy, 15*(6), 650–662. doi: 10.1111/j.1532-7078.2010.00032.x

Schweinberger, S. R., Walther, C., Zäske, R. and Kovács, G. (2011), Neural correlates of adaptation to voice identity. *British Journal of Psychology, 102,* 748–764. doi: 10.1111/j.2044-8295.2011.02048.x

Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804–815.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language & Hearing Research*, *57*(1), 285–296. doi:10.1044/1092-4388(2013/12-0314

Spence, M.J., Rollins, P.R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech Language and Hearing Research, 45*(1), 214–222.

Svoboda, E., McKinnon, M.C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia, 44*, 2189–2208.

Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1,* 121–131.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A.L. (2003). Modulation of neural responses to speech by direction attention to voices or verbal content. *Cognitive Brain Research, 17*, 48–55.

Wong, P.C.M., Nusbaum, H.C., & Small, S.L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16,* 1173–1184.

VITA