

Liu vs. Liu vs. Luke? Name influence on voice recall

BRIANNE SENIOR, JOBIE HUI and MOLLY BABEL
University of British Columbia

Received: February 27, 2017 Revised: March 6, 2018 Accepted: March 17, 2018

ADDRESS FOR CORRESPONDENCE

Molly Babel, 2613 West Mall, Department of Linguistics, University of British Columbia,
Vancouver, BC V6T 1Z4 Canada. E-mail: molly.babel@ubc.ca

ABSTRACT

Listeners are better at remembering voices speaking in familiar languages and accents, and this finding is often dubbed the *language-familiarity effect* (LFE). A potential mechanism behind the LFE relates to a combination of listeners' implicit knowledge about lower level phonetic cues and higher level linguistic processes. While previous work has established that listeners' social expectations influence various aspects of linguistic processing and speech perception, it remains unknown how such expectations might affect talker recognition. To this end, Mandarin-accented English voices and locally accented English voices were used in a talker recognition paradigm in conditions which paired voices with stereotypically congruent names (Mandarin-accented English voice as *Chen* and locally accented English voice as *Connor*) and stereotypically incongruent names (vice versa). Across two experiments, listeners showed greater recall for the familiar, local voices than the Mandarin-accented ones, confirming the basic premise of the LFE. Further, incongruent accent/name pairings negatively affected listeners' performance, although listeners with experience speaking Mandarin were less influenced by the incongruent accent/name pairings. These results indicate that the LFE, while relying largely on listeners' ability to parse linguistic information, is also affected by nonlinguistic information about a talker's social identity.

Speech perception goes beyond extracting information from an unfolding acoustic signal. Listeners must store information from the speech stream for later use for both linguistic and social purposes. For example, when confronted with a speech signal, listeners must retain information about that talker's voice so that when they later encounter the same voice, they are able to identify the talker. We are able to recognize the voice of a familiar talker (e.g., your best friend) regardless of whether they say, for example, *dogs* /dagz/ or *cats* /kæts/, despite the phonemic differences between these words. Previous research has shown that our ability to learn and later recognize and recall voices, here termed *talker recognition*, depends on our familiarity with the language background of the speakers. In general, listeners are better able to identify voices speaking in the listener's native language than they are at identifying voices in a foreign language (e.g., Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007; Thompson, 1987; Winters, Levi, & Pisoni, 2008), and this finding has been dubbed the *language-familiarity effect* (LFE). This proposed effect describes the

observation that, for example, monolingual English-speaking listeners better recognize voices speaking English than they do if the same voices are speaking an unfamiliar language such as German, whereas German monolingual listeners show the reverse pattern of results (Goggin et al., 1991). Similar observations have been made when listeners are exposed to a speaker of their native language who has an unfamiliar or foreign accent. Here, listeners show decreased accuracy in identifying foreign-accented speakers, with recognition accuracy for voices of the listener's "own-accent" surpassing accuracy for voices with a less familiar or foreign accent (e.g., Goggin et al., 1991; Stevenage, Clarke, & McNeill, 2012; Thompson, 1987). These findings would seem congruent with a LFE for accents, where listeners are not only better at recognizing talkers of a familiar language but also familiarly accented talkers.

A proposed mechanism behind the LFE in voice recognition lies in listeners' implicit knowledge about which phonetic cues are talker-specific in a familiar language, but not being able to tease apart talker-specific and language- or accent-general phonetic cues in an unfamiliar language (Perrachione, in press; Winters et al., 2008). That is, listeners have implicit knowledge about the distribution of acoustic-phonetic parameters for sounds in their language (e.g., vowel formants, voice onset time, energy distribution in a fricative, timing relationships between articulatory and laryngeal movements, etc.) and are able to parse phonetic variation accordingly in a familiar language. However, being unfamiliar with the sound categories in an unfamiliar language, listeners do not know whether a phonetic cue indicates a particular sound category or a talker-specific trait. Any model of spoken language recognition relying entirely on the bottom-up intake of the acoustic-phonetic distributions of sounds, however, can only take an experienced listener so far. Listeners use knowledge about words to facilitate the categorization and recognition of the phonetic signal. For example, in the classic Ganong effect, when presented with a sound ambiguous between /t/ and /d/, listeners will be more likely to categorize the sound as /t/ when it is spliced onto /æsk/, forming the real word *task* versus the nonword *dask* (Ganong, 1980). Thus, listeners' perceptions of phonetic variation are heavily influenced by probabilistic phonological and lexical knowledge (e.g., Pitt & McQueen, 1998; Pitt & Szostak, 2012), and there is convincing evidence from selective adaptation that these are not decision-level processes (Samuel, 2001).

In addition to the linguistic influences on perception and recognition, listeners' social expectations can exert their own influence on linguistic processing and perception. As listeners parse the speech stream, they use information from their experiences to generate expectations about speakers and their social identities (for an overview, see Drager, 2010). These expectations are often stereotyped and overgeneralized deterministic views about the connections between accents, speech styles, and sociocultural behaviors. Listener expectations about a speaker's age, socioeconomic status, gender, nationality, and sexual orientation affect linguistic processes such as vowel and fricative categorization (Drager, 2011; Hay, Warren, & Drager, 2006; Johnson, Strand, & D'Imperio, 1999; Munson, Jefferson, & McDonald, 2006; Niedzielski, 1999; Strand & Johnson, 1996). At the lexical level, listener expectations and associations between accent and

race affect recognition of words (Staum Casasanto, 2008). Categorization and recognition at the phoneme and word levels are the building blocks involved in more global assessments like perceived accentedness and intelligibility (quantified as the proportion of words correctly recognized in an utterance). For example, listeners' stereotyped expectations about the relationship between language and ethnicity appear to bias listeners to perceive the speech from ethnically Asian individuals as less intelligible and more accented (Babel & Russell, 2015; Rubin, 1992;; see also Devos & Banaji, 2005). Yi, Phelps, Smiljanic, & Chandrasekaran (2013) suggest that the loss of intelligibility of Korean-accented English is at least partly related to American listeners' beliefs about Asianness and foreignness.

Speech perception is thus a process that leverages both linguistic and social knowledge. If listener expectations influence linguistic processing, is it also the case that listeners' social associations influence talker recognition? In the following experiments, we manipulate listeners' social expectations by manipulating the types of names assigned to various talkers to examine the role of social factors in talker recognition. Typically, talker recognition tasks involve a training phase in which listeners are familiarized with the different voices and an identification phase in which listeners must recognize and identify various talkers. Identification phases have consisted of recognizing old and new talkers in a voice lineup (Goggin et al., 1991; Thompson, 1987), identifying the avatar (Bregman & Creel, 2014; Orena, Theodore, & Polka, 2015; Perrachione, Del Tufo, & Gabrieli, 2011) or number (Xie & Myers, 2015) associated with each voice, and providing the name associated with each voice (Perrachione & Wong, 2007; Winters et al., 2008). In the latter method, each voice is paired with a name, and listeners respond by selecting the name of the talker they think is speaking. Little has been reported on the names used in these studies. For example, Perrachione and Wong (2007, p. 1902) used "language-appropriate, familiar monosyllabic names" that included the Chinese names *Chen, Hong, Liu, Peng, and Wei*. Given the robust evidence in the sociophonetic literature that listeners' social expectations affect categorization of phonemes, words, and global talker properties, it is likely that these methodological decisions may have introduced social information beyond the acoustic signal that could have affected how accurately listeners recalled the voices, as talker recognition tasks rely on listeners learning a speaker's vocal characteristics (i.e., *who* is talking) rather than the linguistic content of what they are saying.

A substantial literature has established that names carry social weight and thus provide information about a speaker. Names elicit different impressions, both positive and negative, about an individual. When exposed to a first name, individuals make assumptions about the person possessing that name, including how caring, successful, trustworthy, and emotionally stable the individual is (Leirer, Hamilton, & Carpenter, 1982; Mehrabian, 2001). Names can also influence how we perceive an individual's physical characteristics, with socially "desirable" names increasing the perceived attractiveness of an individual (Garwood, Cox, Kaplan, Wasserman, & Sulzer, 1980). Such assumptions elicited by names may even lead us to treat others differently solely on the basis of their names. Bertrand

and Mullainathan (2004) found that employers in Boston and Chicago are more likely to respond favorably to resumes with White-sounding names than to the same resumes with stereotypically Black-sounding names. In a similar study, Spriestma (2013) sent essays to primary teachers in Germany and had them score each essay. Some of the writers of these essays were assigned Turkish names (a large, migrant population in Germany) whereas others were given German names; a minority of teachers scored essays associated with a Turkish name significantly lower than similar quality essays paired with German names. Similar findings have been found by researchers in the United States where teachers have been shown to assign lower scores to students with African American-sounding names than to students with White-sounding names, even when the student's race was controlled for (Anderson-Clark, Green, & Henley, 2008).

Taken together, this body of literature establishes that names are important as they interface with individuals' expectations. Given that social expectations affect recognition processes from sounds to sentences, we test whether social expectations also affect talker recognition. We do this in an experiment that pairs natively accented English voices and Mandarin-accented English-speaking voices with English names and Chinese names (e.g., *Luke* vs. *Liu*). Listeners are either presented with congruent accent/name pairs (e.g., a natively accented English speaker assigned the name *Luke* and a Mandarin-accented speaker assigned the name *Liu*) or incongruent accent/name pairs (e.g., a native-accented English speaker named *Liu* and a Mandarin-accented speaker named *Luke*). We test whether names affect voice recognition on listeners with and without experience with the Mandarin language. If the LFE relies largely on listeners' ability to parse phonetic information as talker specific or accent general, we would not expect accent/name congruency to affect listeners' abilities to recall voices; we would only predict that listeners with Mandarin experience will perform more accurately on the Mandarin-accented voices than listeners with no Mandarin knowledge. Given the social implications imbued in names and the influence of social knowledge and listener expectations on the recognition of the speech stream, we predict a reduction in listeners' talker recognition performance when the talkers in the task are assigned a name that is incongruent with the talker's accent. Experiment 1 establishes that accent/name pairings affect talker recognition accuracy in a between-listener design. Experiment 2 replicates this basic finding in a within-listener design and attempts to decipher the mechanism that decreases performance in incongruent accent/name pairings.

EXPERIMENT 1

Method

Speakers. We recorded five speakers of Mandarin-accented English and five native English monolingual speakers. Mandarin speakers self-identified as native speakers of Mandarin Chinese, none reported being fluent speakers of English, and none reported using English as their dominant language. At the time of

recording, speakers had spent an average of 3 years in Vancouver (range = 1–5; $SD = 1.5$). All Mandarin-accented speakers were male and had a mean age of 25 years (range = 19–27; $SD = 3.3$). Native English speakers reported speaking a variety of Canadian English, had spent their entire life on the West Coast of British Columbia (either in the Lower Mainland or on Vancouver Island), and had lived in Vancouver for at least 5 years. All speakers were male, with a mean age of 23 years (range = 21–27; $SD = 2.5$).

The two sets of voices were collected separately. The native English speakers were recorded using a Sound Devices USB PreAMP and head-mounted microphone with a 44,100-Hz sampling rate. Speakers read 10 sentences taken from Bradlow and Alexander (2007). All were declarative sentences in either the present ($n = 6$) or past tense ($n = 4$) with one or two clauses. Five of these sentences consisted of a final word that was highly predictable from the preceding context while the other 5 sentences had a final word that was unpredictable. The high-predictability sentences were slightly longer than the low-predictability sentences. The high-predictability sentences contained five to nine words ($M = 7$, $SD = 1.58$), whereas the low-predictability sentences all contained five words. A full list of sentences is provided in Appendix A. Recording took place in a sound-attenuated booth. Each sentence was presented on the screen using Microsoft PowerPoint, and participants advanced to the next slide when they finished reading the sentence. They were asked to read each sentence three times, and the best utterance, that which was free of false starts, fillers, and mispronunciations, was selected. The Mandarin speakers were recorded with a standing microphone and Sound Devices USB PreAMP in a quiet room.¹ Due to a technical error, three of the Mandarin-accented speakers were recorded with a 48,000-Hz sampling rate while the remaining speakers were recorded at 44,100 Hz. The Mandarin speakers read all 120 sentences from Bradlow and Alexander (2007), but only the 10 aforementioned sentences that were recorded by the native English speakers were used. This ensured that the sentence lists were identical across the two voice sets. An experimenter was present during recording and asked the speaker to repeat the sentence if he or she made any mistakes or did not produce the sentence fluently.

Sentences from all 10 speakers were trimmed to remove leading and trailing silence in Praat (Boersma & Weenink, 2017); all trimming was done at zero crossings to avoid introducing transients into the recordings. On average, recordings from the Mandarin speakers were longer and more variable in length ($M = 2.25$ s, $SD = 1.56$) than those from the native English speakers ($M = 1.56$ s, $SD = 0.35$), and within each set of voices, high-predictability sentences were longer (Mandarin-accented: $M = 2.59$ s, $SD = 0.65$; native English: $M = 1.74$ s, $SD = 0.40$) than low-predictability sentences (Mandarin-accented: $M = 1.91$ s, $SD = 0.36$ seconds; native English: $M = 1.37$ s, $SD = 0.14$). All files were RMS amplitude normalized to 73 dB in Praat, and in this process it was confirmed that no samples were clipped.

If a talker's voice is highly variable, it may be harder to consistently learn and associate with a name. However, if all of the voices from an accent together make the accent more variable, that variability may aid in making each voice within an

accent perceptually unique. To quantify accent variability, Table 1 provides summary statistics for selected acoustic measures by accent. Table 2 provides these data by talker. Because all speakers produced the same sentences, the duration of each sentence is equivalent to speech rate. Fundamental frequency was estimated using REAPER (<https://github.com/google/REAPER>). The lower and upper 10% quantiles were removed to account for measurement misestimation, and the remaining data were standardized to semitones on a per-talker basis. Variance ratios were used to compare these average values for duration and normalized f0 for the two accents (following Johnson, Westrek, Nazzi, & Cutler, 2011). The average duration and normalized f0 values for the Mandarin-accented voices tended to be more variable as a group in terms of duration, Mandarin-accented: $M=2.25$ s, $SD=0.32$, native English: $M=1.56$ s, $SD=0.12$; $F(4, 4)=6.53$, $p=.105$, and normalized f0, Mandarin-accented $M=3.29$ semitones, $SD=1.22$, native English $M=2.58$, $SD=0.41$; $F(4, 4)=8.78$, $p=.03$.² These trends for the Mandarin-accented English speakers to be more variable than the native accent English speakers would potentially assist listeners in associating and identifying these Mandarin-accented voices, as each voice would potentially be more perceptually unique.

We also conducted an analysis of voice similarity for two reasons: to establish that the voices merit being categorized as two accents and to determine that one set of voices is not inherently more similar than the other on a spectral level. While an ideal analysis of accent and speaker similarity may be composed of listener judgments (though sensitivity to such parameters are a side effect of the LFE), we compared speakers acoustically using mel-frequency cepstral coefficients (MFCCs, a method of speech parameterization frequently used for speech recognition) and a dynamic time warping algorithm using Phonological Corpus Tools (Hall, Allen, Fry, Mackie, & McAuliffe, 2016). The MFCC analysis used 26 triangular filters and 12 orthogonal coefficients. The dynamic time warping algorithm sought the lowest cost path through a distance matrix that was time independent. The distance value was output as a similarity value, where 1 equals a perfect match. All matched sentences were compared across 10 speakers such that there were 100 native English to native English comparisons, 100 Mandarin to Mandarin comparisons, and 450 Mandarin to English comparisons. The cross-accent voice comparisons were more different

Table 1. *Selected summary acoustics by accent*

Accent	Duration in seconds (all sentences)	Duration in seconds (high-predictability sentences)	Duration in seconds (low-predictability sentences)	f0 in semitones
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Mandarin	2.25 (1.56)	2.59 (0.65)	1.91 (0.36)	3.28 (0.58)
Native English	1.56 (0.35)	1.74 (0.40)	1.37 (0.14)	2.57 (0.36)

Table 2. Selected summary acoustics by talker

Talker	Accent	Duration in seconds Mean (SD)	f0 in semitones	
			Mean (SD)	Min, Max
CHM109	Mandarin	2.53 (0.61)	2.36 (1.31)	0, 5.2
CHM111	Mandarin	1.74 (0.53)	2.4 (1.42)	0, 5.87
CHM113	Mandarin	2.30 (0.51)	5.14 (2.63)	0, 9.81
CHM116	Mandarin	2.19 (0.56)	3.94 (2.25)	0, 8.63
CHM117*	Mandarin	2.48 (0.65)	2.59 (1.53)	0, 6.01
ENM101*	Native English	1.44 (0.21)	2.58 (1.56)	0, 5.68
ENM102	Native English	1.48 (0.31)	1.91 (1.05)	0, 4.02
ENM103	Native English	1.56 (0.32)	3.02 (1.97)	0, 7.05
ENM104	Native English	1.55 (0.33)	2.59 (1.8)	0, 6.62
ENM105	Native English	1.76 (0.49)	2.78 (1.8)	0, 6.66

Note: Talkers with * were used only in Experiment 1. ENM, native English-speaking males. CHM, Chinese males speaking Mandarin-accented English.

($M=0.02$, $SD=0.001$) than either within-accent voice comparison (English–English pairs: $M=0.023$, $SD=0.002$; Mandarin–Mandarin pairs: $M=0.024$, $SD=0.002$). Mean by-talker combination (averaged over the 10 sentences) similarity values were used as the dependent measure in analysis of variance (ANOVA) with accent (Mandarin–English, English–English, Mandarin–Mandarin) as the independent variable, returning a main effect of accent, $F(2, 42)=26.88$, $p<.001$. Planned comparisons confirm that the Mandarin–English comparisons were more different than Mandarin–Mandarin, $t(11)=5$, $p=.0004$, and English–English, $t(11.5)=5.18$, $p=.0003$, comparisons, and that the two same-accent comparisons are not different, $t(17.75)=-0.28$, $p=.78$. This confirms that the within accent voices do not differ in terms of acoustic-phonetic variability. To confirm that the voices differed acoustically in terms of perceivable accent, the similarity values were translated into distance values, which were multidimensionally scaled to two dimensions. This computation is shown in Figure 1. Dimension 1 provides a clear separation between the Mandarin-accented voices and the native English voices, indicating that on acoustic-phonetic grounds, these voices are easily separable into distinct groups. Dimension 2 shows vertical spread for each of the accent groups, with a clear voice outlier in CHM111 for the Mandarin accented group. This suggests that on acoustic-phonetic grounds, the Mandarin accented voices might be more distinct because of the outlier status of CHM111, which might give these voices an advantage in the talker recognition task (Kreiman & Papcun, 1991; Orchard & Yarmey, 1995; Papcun, Kreiman, & Davis, 1989, Yarmey, 1991). It could also be the case that Dimension 2 accounts for residual variance that is neither interesting nor meaningful.

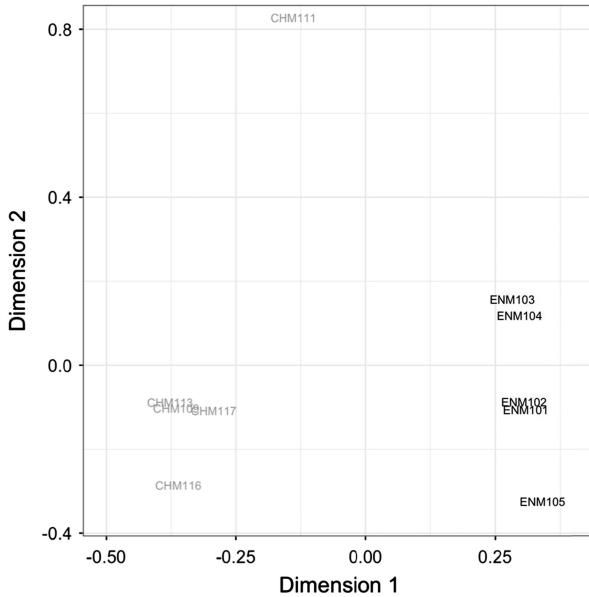


Figure 1. Two-dimensional scaling solution based on the mel-frequency cepstral coefficients (MFCC) acoustic-similarity of the 10 voices. Labels beginning with ENM are native English-speaking males, and those with CHM are Chinese males who speak Mandarin-accented English.

Names. We used five stereotypically North American (“English”) male names (*Connor, Gabriel, John, Luke, and Steven*) and five Romanized Chinese male names (*Chen, Hong, Liu, Peng, and Wei*). Within each subset of names, we avoided names that shared the same onset consonant to alleviate any effects of phonological competition.

We generated a list of common North American English names based on the experimenters’ intuitions and experiences with names in Vancouver, British Columbia. These names were pretested with a name/ethnicity association pretest in which participants were asked for each name “Which ethnicity do you most associated this name with?” Possible responses were White Canadian, Asian Canadian, and other. On the basis of these preliminary results, we selected five potential English names: *Bob, Connor, Jake, Luke, and Stephen*. These names were then compared to one another using statistics from the British Columbia Vital Statistics Agency (<http://www.vs.gov.bc.ca/babynames/>). We sought to select English names with similar birth popularity trajectories in British Columbia over the past 20 years. Because *Bob* did not exist as a name in the database and the similar name *Bobby* was comparably less popular, this name was swapped with the name *Gabriel*; this name better matched the popularity trajectory of the other English names. For the same reason, *Jake* was replaced with *John*, and the spelling of *Stephen* was changed to *Steven*.

As the Vital Statistics Agency only reports names for which five or more babies are born with that name, the relative infrequency of common Chinese baby names prevented us from retrieving similar population trajectories in choosing Chinese names. When selecting Chinese names, we used the names from Perrachione and Wong (2007). We favored these names for reasons relating to phonetic composition and cultural naming practices. These names did not include any phonemes that are not in the English phonetic inventory, such as the onsets in names like Xiu [çiou] and Qing [tɕ^hiŋ], which may have caused additional difficulty for participants unfamiliar with Mandarin phonology. Names that are harder to pronounce (e.g., *Colquhoun*) are dispreferred compared to names that are less difficult to pronounce (e.g., *Smith*), even when other factors such as name length and orthographic regularity are controlled for (e.g., Laham, Koval, & Alter, 2012).

To confirm intuitions about the language backgrounds associated with each name, we conducted a visual analogue scale-rating task to assess the names. Twenty-one individuals from the University of British Columbia community voluntarily completed this task after they had taken part in an unrelated study. Participants were presented with a form in hard-copy that asked “Who do you associated the name _____ with?” and then instructed to make a mark in pen along a 10-cm scale with endpoints “native speaker of Mandarin” and “native speaker of English” to indicate their rating. A vertical dash indicated the midpoint on the scale. We controlled for left-right positivity associations by giving half the participants scales with “native speaker of Mandarin” as the left endpoint and half with the same label as the right endpoint. The names were displayed alphabetically to avoid ordering them in a way that might influence ratings.

We converted these ratings into a numerical scale, coding a response of “native speaker of English” as 100 and a response of “native speaker of Mandarin” as 0. Three participants chose to circle the endpoint label; in such cases, their response was counted as either 100 or 0 depending on the label circled. All other measurements were taken to the nearest millimeter. While most participants marked a vertical line on the scale to indicate their rating, other participants responded by marking Xs, circles, upside-down triangles, and checkmarks along the scale. Xs were measured from the intersection of the two diagonal lines, and circles (or ovals) were measured from the midpoint of the circle’s horizontal diameter. Both upside-down triangles and checkmarks were measured from the bottom vertex. Two participants each had one name where they marked their response ambiguously (i.e., they made multiple marks on the scale), and these two data points (one name scale for each participant) were discarded.

The five names divided into the two name groups are shown in Figure 2 and confirm our intuitions about which type of name was associated with which type of speaker in Vancouver. As a group, our Chinese names (*Chen, Hong, Liu, Peng, and Wei*) received lower ratings ($M = 15, SD = 3$) than the North American English names (*Connor, Gabriel, John, Luke, and Steven; M = 83, SD = 3*). Note that while both types of names occupy the middle range of the continuum, the distribution peaks higher in the middle for the more stereotypically English

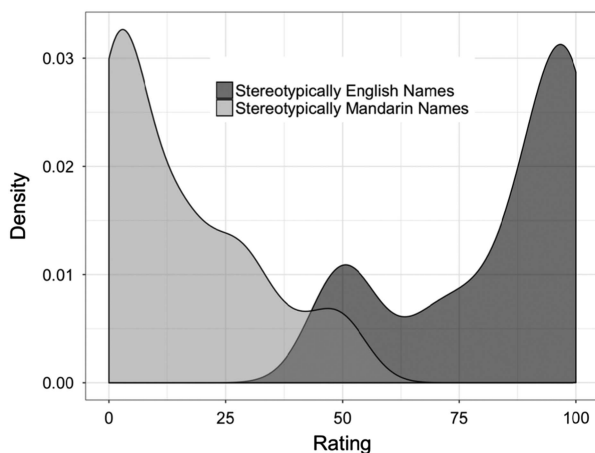


Figure 2. Density plot of the ratings of the names used in Experiments 1 and 2 where native speaker of Mandarin (0) and native speaker of English (100) were the endpoints.

names. This suggests that some participants view these names as equally appropriate for native speakers of English or Mandarin.

To test whether the Chinese names were more challenging to learn and associate with entities generally, we conducted a brief experiment in the same format (exposure, training, test) as the talker recognition task. Instead of voices, the names were paired with simple shapes selected from the preset shape options in Adobe Photoshop. Each shape was filled with a unique color. A small sample ($n = 14$) of participants completed the task where name type (Chinese or English) was blocked and counterbalanced across participants. To account for the fact that some shapes may be much more unique than others, name/shape pairings were randomized across participants, with each participant receiving a different set of pairings. Participants' performance varied considerably on this task, but they were overall very accurate. Performance at test was insignificantly more accurate on the English names ($M = 92\%$, $SD = 18$) than the Chinese names ($M = 87\%$, $SD = 23$), $t(13) = 1.02$, $p = .33$, Cohen's $d = 0.24$. This suggests that with greater power, the small benefit for the English names may have been reliable, but there is no large difference between how easily learned and associated these names are.

Listeners. The recruitment, compensation, and procedures for this study were approved by the Behavioural Research Ethics Board at the University of British Columbia. One hundred twenty-nine English-speaking listeners completed the talker recognition task. They were recruited through the University of British Columbia linguistics subject pool and were compensated with either course credit or \$7 for their time. One participant was removed because of a self-reported hearing disability. This left 128 participants. Of these participants, 65 self-reported experience speaking

Mandarin (mean age of 20.5, range = 17–48, *SD* = 4.1), whereas the remaining 63 participants had no experience with Mandarin (mean age of 21.6, range = 17–45, *SD* = 5.6).

Of the 65 Mandarin-speaking participants, 25 reported speaking Mandarin fluently, 20 reported speaking fairly well, and 20 reported speaking poorly. Of these participants, 29 reported also speaking English natively and a further 21 said English was the dominant language they used. Fifteen participants did not report English as their dominant language. On average, these participants started learning Mandarin at 5.8 years of age (median age of 4 years, range = 0–25, *SD* = 5.9).³ Within the non-Mandarin-speaking group, 48 of the 63 participants reported English as their native language, 9 reported English as their dominant language, 5 did not have English as their native or dominant language, and 1 did not specify. While we base our categorization of participants based on their self-report of speaking Mandarin, we refer to this distinction simply as Mandarin experience, generalizing their self-reported speaking abilities to assume listening experience with Mandarin as well. Tables 3a and 3b provide a summary of the listener groups for Experiment 1.

Procedure

Our methods followed Perrachione and Wong (2007). Participants completed a talker recognition task consisting of five parts: (a) exposure to voices, (b) training, (c) practice quiz, (d) final test, and (e) a questionnaire. All parts were presented using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA), and listeners wore AKG K240 headphones while seated in sound-attenuated cubicles. Participants received verbal instructions at the beginning of the experiment and

Table 3. *Listener demographics (Experiment 1)*

<i>a. Listeners' self-reported experience speaking Mandarin for Experiment 1</i>		
Total number of participants	128	
Self-reported experience with speaking Mandarin	65	
Fluent	25	
Fairly well	20	
Poorly	20	
Self-reported no experience with speaking Mandarin	63	
<i>b. Listeners' self-reported familiarity and use of English for Experiment 1</i>		
	Mandarin-speaking participants	Non-Mandarin-speaking participants
English used as native language(s)	29	48
English used as dominant language	21	9
English neither native nor dominant language	15	5
No specification for use of English	0	1

written instructions on screen prior to each part. Following Perrachione and Wong (2007), we emphasized that the subjects should focus on who was talking rather than the content of the speakers' utterances. The experiment took approximately 30 min to complete.

Exposure. Listeners were first familiarized with the five voices and their associated names. They saw the first name on the screen and heard the voice associated with that name produce a sentence. Then, they saw the next name appear on the screen and heard the next speaker produce the same sentence. This continued until listeners had been introduced to all five speakers and their assigned names. In all conditions, the exposure sentence was always the high-predictability sentence *The good boy is helping his mother and father.*

Training. The training phase consisted of five trials, one corresponding to each of the five speakers. On each trial, listeners were presented with the same sentence they had heard in exposure; this sentence was read by one of the five speakers. Participants saw all five names on the screen (in a random order that remained constant across all participants), and they were asked to identify the name of the speaker who was reading the sentence by pressing the corresponding button on the response box. Participants received feedback on whether their response was correct; if incorrect, they were provided with the name of the correct speaker. Participants had 5000 ms to respond. If they did not respond within that interval, participants received a message on the screen informing them that no response had been detected, and the experiment advanced to the next trial.

Practice quiz. Listeners performed the same task as in the training phase. However, they were presented with four additional, novel sentences and asked to generalize what they had learned about the voices to novel stimuli from the same speakers. The practice quiz consisted of 25 trials (5 sentences \times 5 speakers), and listeners received feedback as to the correct response.

Final test. In the final test, listeners performed the same task, but had to identify 5 new sentences in addition to the sentences they had been exposed to in the previous three sections. The final test consisted of 50 trials (10 sentences \times 5 speakers), and participants did not receive feedback about whether their response was correct.

Questionnaire. To close, participants completed a short questionnaire evaluating their language background and experience speaking Mandarin. A copy of this questionnaire is available in Appendix B.

Excluding the questionnaire, participants underwent this process twice, with one block for each accent type. Upon completion of the two blocks, participants responded to the questionnaire. Accent was a within-subjects factor and both accent/name congruency and Mandarin experience were between-subjects factors.

Results

Trials without a response ($n=165$) were removed from the data set. The remaining data were analyzed in a repeated measure ANOVA. Speaker accent (Mandarin, native) was repeated across listeners, and accent/name pairing (congruent, incongruent) and Mandarin experience (Mandarin–English bilingual, no Mandarin) were between-listener dependent variables. The dependent variable was listener accuracy averaged across trials.

There was a main effect of speaker accent, $F(1, 124)=174.27, p<.001, \hat{\eta}_G^2=0.29$, which replicates the LFE on an accent level. Listeners were more accurate on the native English-speaking voices ($M=65\%, SD=9$) than the Mandarin-accented voices ($M=45\%, SD=3$). There was also a main effect of accent/name pairing, $F(1, 124)=5.43, p=.021, \hat{\eta}_G^2=0.03$. Listeners were more accurate when accent/name pairings were congruent ($M=58\%, SD=3$) than when they were incongruent ($M=52\%, SD=3$). While listeners who were Mandarin–English bilinguals performed more accurately on the task ($M=58\%, SD=3$), overall compared to listeners with no Mandarin experience ($M=53\%, SD=3$), this effect was not significant, $F(1, 124)=3.47, p=.07, \hat{\eta}_G^2=0.0196$. None of the interactions were significant: Mandarin Experience \times Accent/Name Pairing, $F(1, 124)=0.01, p=.92, \hat{\eta}_G^2=0.00006$; Mandarin Experience \times Accent: $F(1, 124)=0.12, p=.73, \hat{\eta}_G^2=0.0003$; Accent \times Accent/Name Pairing: $F(1, 124)=1.51, p=.22, \hat{\eta}_G^2=0.003$; Mandarin Experience \times Accent \times Accent/Name Pairing: $F(1, 124)=2.17, p=.143, \hat{\eta}_G^2=0.005$.

Figure 3 presents accuracy data from Experiment 1 with the y-axis showing the proportion of correct responses for the Mandarin accented voices and native accented voices for the congruent (light gray bars) and incongruent (darker gray bars) accent/name pairs. The x-axis separates these data according to listeners' experience with Mandarin. Listeners were more accurately overall with native accents, and listeners were overall more accurate with congruent accent/name

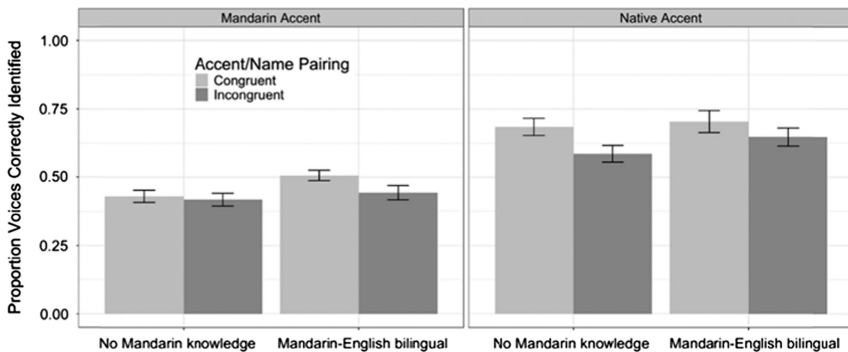


Figure 3. Proportion of voices correctly identified (y-axis) by name in Experiment 1 for (a) Mandarin accented voices and (b) native English voices. The x-axis separates listeners based on whether they are Mandarin–English bilinguals. Lighter gray bars show congruent accent/name pairings, and the darker gray bars show incongruent accent/name pairings. Error bars show standard error.

pairings. Mandarin–English bilinguals tended to be more accurate at learning the names and the voices, but this appears to be somewhat variable across the accent and name combinations.

Discussion

The results of this experiment partially replicate the basic premise of the LFE: listeners were overall more accurate at remembering speakers' names and faster at providing those responses in a familiar accent, the local accent, than in a less familiar accent, Mandarin-accented English. Despite our acoustic measurements indicating that the Mandarin voices may have been slightly more dissimilar and thus more perceptually unique than the native English voices, listeners did not find it easier to identify the Mandarin-accented voices. However, the Mandarin–English bilinguals did not perform better on the Mandarin-accented voices than the group without Mandarin experience as the LFE would predict. It is possible that despite not speaking Mandarin, the non-Mandarin-speaking listeners still had enough relevant experience with hearing Mandarin-accented English due to the prevalence of Mandarin speakers in Vancouver.

Listeners who were Mandarin–English bilinguals tended to do better overall on the task compared to the listeners who had no experience with Mandarin, but not significantly so. This runs counter to previous findings such as that of Xie and Myers (2015), who found that Mandarin–English bilinguals outperform monolingual English listeners on talker recognition tasks for both familiar and unfamiliar accents. Xie and Myers argue that listeners with experience with tone perception, through either musical training or experience speaking a tone language like Mandarin, show improved talker recognition accuracy.⁴ Our Mandarin–English bilingual listener populations differ from other Mandarin–English populations studied within the talker recognition literature (e.g., Xie & Myers, 2015, who used late English learning bilinguals). In Vancouver, it is not the case that most Mandarin speakers, even those who speak Mandarin natively, learned English later in life. The 22 native Mandarin speakers collectively have an average age of English onset of 8.42 (median = 8, $SD = 3.46$). Over half of these 22 native Mandarin speakers reported also speaking English fluently, and all but 1 speak English fluently or fairly well. Many Mandarin speakers in Vancouver are simultaneous Mandarin–English bilinguals, Mandarin heritage speakers, or early second language English learners. In addition, the balance of our Mandarin–English population skews toward English dominance; 77% of the Mandarin–English bilinguals in this sample were also native speakers of English or identified English as their dominant language. This tipping of the scales toward English could indicate that this sample did not have enough experience with tones to show a global boost in talker recognition, as predicted by Xie and Myers. This all builds to the Mandarin–English bilinguals in our study *not* being necessarily more familiar with Mandarin-accented English than the local native variety of English. Therefore, it is perhaps not surprising that Mandarin–English bilinguals in our sample would perform better on the native English voices than the Mandarin-accented voices. Rather, since the Mandarin–English listeners are more

familiar with Mandarin-accented English than the listeners with no Mandarin experience, we might have expected the Mandarin–English bilinguals to do better on the Mandarin-accented English voices than the listeners without any Mandarin experience; however, we found no evidence of this interaction.

It merits mention that the demographics of Vancouver make our listener samples for both groups extremely heterogeneous, consisting of both self-reported native and nonnative English speakers. Crucially, this means that any differences between the groups are presumably not solely due to differences in native English speaker status.

Performance in our task was crucially affected by social cues in the form of accent/name assignment. Listener performance was more adversely affected by incongruent accent/name pairings. This suggests that in addition to talker recognition relying on a listener's familiarity with the accent or language, it is also affected by perceived fit of the speakers' names. There are two possible interpretations for the finding that incongruent accent/name pairings negatively affect native accents more than Mandarin ones. One possibility is that the pairing of a voice and name affects the perceived accentedness of the voice (Babel & Russell, 2015; Kang & Rubin, 2009), causing an illusory lack of familiarity and attenuation of the LFE or a perceived increase in familiarity and enhancement of the LFE, though, to our knowledge, there is no evidence for illusory effects of the LFE on talker recognition. A second possible interpretation is that listeners assume speakers have stereotypically Chinese or English names, and thus have a more challenging time making unexpected accent/name associations.

Experiment 2 tests this latter possible interpretation by using a new design that pairs Mandarin and native accents with stereotypically Chinese and English names in single blocks that are either congruent or incongruent accent/name pairings. As in Experiment 1, accent was a within-subjects factor; crucially, in Experiment 2, accent/name congruency was also made a within-subjects factor. In Experiment 1, listeners completed two blocks, one for each set of voices, and were presented with names that were either always congruent with the speaker's accent or always incongruent with the speaker's accent. In Experiment 2, listeners again completed two blocks. Here, however, each block contained both native English voices and Mandarin-accented voices, and the first block contained congruent accent/name pairings while the second contained incongruent accent/name pairings or vice versa. This design permits an error analysis to see whether the name type affects the kind of errors made. Upon completion of the talker recognition task, participants in Experiment 2 also completed an accentedness rating task with or without names to examine whether name association colors the perceived foreign accentedness of speakers' voices.

EXPERIMENT 2

Experiment 2 endeavored to elucidate the mechanism behind the results of Experiment 1 by using a design that combined Mandarin and native accents in a single block.

Method

Speakers. The same speakers as in Experiment 1 were used. To alleviate the overall difficulty of the task, we present listeners with four voices per block (two Mandarin-accented speakers and two natively accented speakers), removing two speakers from the design. The most accurately identified Mandarin-accented speaker and the least accurately identified native accented speaker were removed from the voice set.

Names. We used the same names as in Experiment 1, but removed the names associated with the removed speakers (*Steven* and *Wei*).

Listeners. The recruitment, compensation, and procedures for this study were approved by the Behavioural Research Ethics Board at the University of British Columbia. Ninety-two listeners completed the task. They were recruited through the University of British Columbia linguistics subject pool and were compensated with either course credit or \$7 for their time. Two participants were removed because they self-reported having speech, language, or hearing disabilities or disorders. This left 90 participants, 46 of whom self-reported experience speaking Mandarin (mean age of 21.3, range = 17–31, $SD = 3$) and 44 of whom had no experience speaking Mandarin (mean age of 24.2, range = 17–66, $SD = 10$).

Of the Mandarin-speaking participants, 20 of these 46 participants also spoke English natively, 8 did not speak English natively but reported it as their dominant language, 17 did not report English as their native or dominant language, and 1 did not specify. On average, Mandarin-speaking participants started learning Mandarin at age 4.1 (median age of 2, range = 0–19, $SD = 5$).⁵ Twenty-eight of these participants reported speaking Mandarin fluently, 9 spoke Mandarin fairly well, and a further 9 reported speaking Mandarin poorly. Within the non-Mandarin-speaking group, 33 of these participants were native English speakers, 6 were not native English speakers but reported English as their dominant language, and 5 did not report English as their native language or their dominant language. Tables 4a and 4b provide listener demographics for Experiment 2.

Procedure

Experiment 2 consisted of two parts: (a) a talker recognition task and (b) a foreign-accentedness rating task. The talker recognition task followed the same structure as in Experiment 1 with the following exceptions: each block contained only four speakers rather than five; within each block, listeners encountered voices of both accent types (i.e., two Mandarin-accented speakers and two native English speakers); and we used a within-subjects accent/name congruency manipulation, with listeners completing both an incongruent and a congruent block. All names were displayed on the screen in alphabetical order. Twelve versions of the experiment were created; in 6, participants completed the congruent block followed by the incongruent block, while in the other 6 the order of

Table 4. *Listener demographics (Experiment 2)*

a. <i>Listeners' self-reported experience speaking Mandarin for Experiment 2</i>		
Total number of participants	92	
Self-reported experience with speaking Mandarin	46	
Fluent	28	
Fairly Well	9	
Poorly	9	
Self-reported no experience with speaking Mandarin	44	
Removed for self-reported speech, language, hearing disorders	2	
b. <i>Listeners' self-reported familiarity and use of English for Experiment 2</i>		
	Mandarin-speaking participants	Non-Mandarin-speaking participants
English used as native language(s)	20	33
English used as dominant language	8	6
English neither native nor dominant language	17	5
No specification for use of English	1	0

these blocks was reversed. The speaker combinations were balanced so that each voice was assigned to both the congruent accent/name condition and the incongruent accent/name condition (i.e., the first Mandarin speaker was called *Liu* in one version of the experiment but *Luke* in another) and so that each speaker appeared with every other speaker at least once across the 12 lists. Across these 12 versions, each speaker was only ever associated with two names (one congruent and one incongruent).

In the second part of the experiment, we exposed listeners to the sentence stimuli from Part 1 and had them make judgments about speaker accentedness. On each trial, participants were presented with a visual analogue scale with endpoints “strong foreign accent” and “no foreign accent” and were instructed to click along the line to make their rating. We controlled for left-right positivity associations by giving half the participants scales with “no foreign accent” as the left endpoint and half with the same label as the right endpoint. While listening to the voice and making their rating, participants saw either a congruent name, an incongruent name, or no name above the scale. The task was blocked by sentence such that listeners heard a single sentence from all 8 speakers before moving onto the next sentence (10 sentences × 8 speakers = 80 trials). This was done to present listeners with the same words and phonological categories to offer a better comparison of pronunciation differences in judging accentedness. Within each sentence block the voices were blocked by accent, with the voices randomized within each accent. The order of the accents within each sentence was counter-balanced between listeners. Sentence order was randomized across participant.

Results

Accuracy. Trials without responses were removed from the data set. For the remaining data set, listener accuracy was used as the independent variable in an ANOVA with speaker accent (Mandarin accent, native accent) and accent/name pairing (congruent, incongruent) as independent variables repeated across listeners and Mandarin experience (Mandarin–English bilingual, no Mandarin experience) as between-listener independent variable.

There was a main effect of speaker accent, $F(1, 88) = 76.87, p < .001, \hat{\eta}_G^2 = 0.12$. Listeners were more accurate at recalling the names of native accented speakers ($M = 86\%, SD = 15$) than Mandarin-accented speakers ($M = 73\%, SD = 9$). There were also main effects of Mandarin experience, $F(1, 88) = 6.37, p = .01, \hat{\eta}_G^2 = 0.03$, and accent/name pairing, $F(1, 88) = 6.87, p = .01, \hat{\eta}_G^2 = 0.01$, as well as an interaction between the two, $F(1, 88) = 4.33, p = .04, \hat{\eta}_G^2 = 0.008$. While Mandarin–English bilinguals ($M = 83\%, SD = 10$) were overall more accurate than listeners with no Mandarin language experience ($M = 76\%, SD = 9.8$) and all listeners were generally more accurate at congruent accent/name pairings ($M = 82\%, SD = 13$) than incongruent accent/name pairings ($M = 78\%, SD = 11$), planned comparisons between congruent and incongruent accent/name pairings for the two listener groups showed that listeners with no Mandarin language experience were more accurate on congruent accent/name pairings, $t(87) = 3.13, p = .002$, Cohen's $d = 0.36$. Mandarin–English bilinguals showed statistically equivalent performance with both congruent and incongruent accent/name pairings, $t(91) = 0.41, p = .68$, Cohen's $d = 0.05$. None of the other interactions in the ANOVA were significant, Mandarin Experience \times Accent: $F(1, 88) = 2.5, p = .12, \hat{\eta}_G^2 = 0.004$; Accent \times Accent/Name Pairing: $F(1, 88) = 0.83, p = .37, \hat{\eta}_G^2 = 0.002$; Mandarin Experience \times Accent \times Accent/Name Pairing: $F(1, 88) = 0.58, p = .45, \hat{\eta}_G^2 = 0.001$.

These data are visualized in [Figure 4](#). Proportion of voices correctly identified at test is shown on the y-axis, and the figure separates performance by listeners' experience with Mandarin (x-axis), accent/name pairing (congruent pairings in light gray, incongruent pairings in dark gray), and speaker accent separated. While Mandarin–English bilinguals are more accurate on the native accent, they show no effect of accent/name pairing. Listeners without any Mandarin language knowledge, however, make more errors on Mandarin and native English accents when the accent/name pairing is incongruent.

Error Type. Within a block, we wanted to count what kind of errors listeners made; that is, when they made errors, did they confuse a speaker with the other speaker of the same accent/name type or did they apply the wrong accent/name entirely? To address this question, we categorized listener errors as either different-accent errors or same-accent errors. There were a total of 1,438 wrong answers across all listeners, which we coded by accent (Mandarin, native), accent/name pairing (congruent, incongruent), and Mandarin experience (Mandarin–English bilingual, no Mandarin knowledge). For all groups, same-accent errors were more common than different-accent errors. That is, listeners were

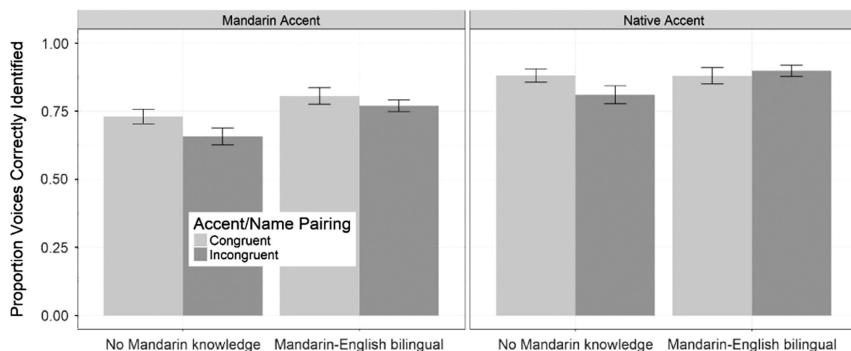


Figure 4. Proportion of voices correctly identified (y-axis) by name in Experiment 2 for (a) Mandarin accented voices and (b) native English voices. The x-axis separates listeners based on whether they are Mandarin–English bilinguals. Lighter gray bars show congruent accent/name pairings, and the darker gray bars show incongruent accent/name pairings. Error bars show standard error.

Table 5. Analysis of deviance table for the generalized regression model for the error analysis

Term	df	Deviance	Residual df	Residual deviance	p value
Accent	1	5.84	1436	1207.4	.016*
Accent/name pairing	1	1.03	1435	1206.4	.31
Mandarin experience	1	5.69	1434	1200.7	.017*
Accent × Accent/Name Pairing	1	2.72	1433	1198	.099
Accent × Mandarin Experience	1	2.31	1432	1195.7	.128
Accent/Name Pairing × Mandarin Experience	1	12.88	1431	1182.8	.0003***
Accent × Accent/Name Pairing × Mandarin Experience	1	1.48	1430	1181.3	.224

Note: p values marked by * indicate significance at the < .05 level, and that marked with *** indicates a p value of < .001.

more likely to confuse a speaker with a speaker of the same accent, regardless of the accent/name pairing or listeners’ backgrounds. However, the degree of this same-accent error bias varied across groups. We tested whether same-accent errors (coded as 1) and different accent errors (coded as 0) were equally common across experimental groups and conditions using a logistic regression with accent error as the dependent variable and accent, accent/name pairing, and Mandarin experience as independent variables. These results are summarized in Table 5 in an analysis of deviance table for the model.

There were more same-accent errors for the Mandarin-accented English voices, demonstrating that regardless of whether a Mandarin-accented voice was named *Chen* or *Connor*, listeners were more likely to confuse it with the other Mandarin-

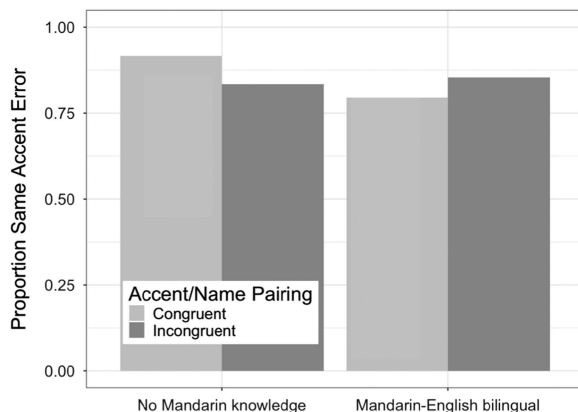


Figure 5. Proportion of same-accent errors in Mandarin Experience (*x*-axis) and accent/name pairing. Congruent accent/name pairings are shown with light gray bars, and the incongruent accent/name pairings are in the darker gray bars. The *y*-axis shows the proportion of same-accent errors.

accented voice in the block. For the natively accented voices, same-accent errors were also more common than different-accent errors, but there were proportionally fewer for the native accents. This demonstrates that the accents of the Mandarin-accented speakers were more confusable, independent of the accent/name pairings.

Mandarin–English bilinguals made fewer same-accent errors than listeners without any knowledge of Mandarin. Language experience also interacted with accent/name pairings, and this interaction is shown in Figure 5. Listeners with no Mandarin knowledge made proportionally more same-accent errors on congruent trials than incongruent trials. This means that a Mandarin-accented voice, whose name was *Liu*, was more likely to be erroneously identified as *Peng* than a Mandarin-accented voice, whose name was *Luke*, was erroneously identified as *Gabriel* or vice versa. Confusing voices of the same accent occurred more when the accent/name pairs were congruent than incongruent for the listeners with no Mandarin knowledge. Mandarin–English bilinguals, by contrast, showed lower same-accent error rates overall, but their performance by accent/name pairings is reversed: they have higher same accent error rates when the accent/name pairings are incongruent. This means that Mandarin–English bilingual listeners made more same-accent errors, proportionally, on blocks with incongruent accent/name pairings. They were more likely to call a Mandarin-accented voice name *Luke*, *Gabriel* than they were to call a Mandarin-accented voice named *Liu*, *Peng*.

Foreign accentedness ratings. Listeners' mouse click locations along the *x*-axis of a line marked with a midpoint and labeled endpoints strong foreign accent and no foreign accent comprise the data for the foreign accentedness ratings. These ratings were normalized by taking the pixel value of the midpoint of the visual analogue line as the scale's midpoint. Scales were flipped such that the left end of

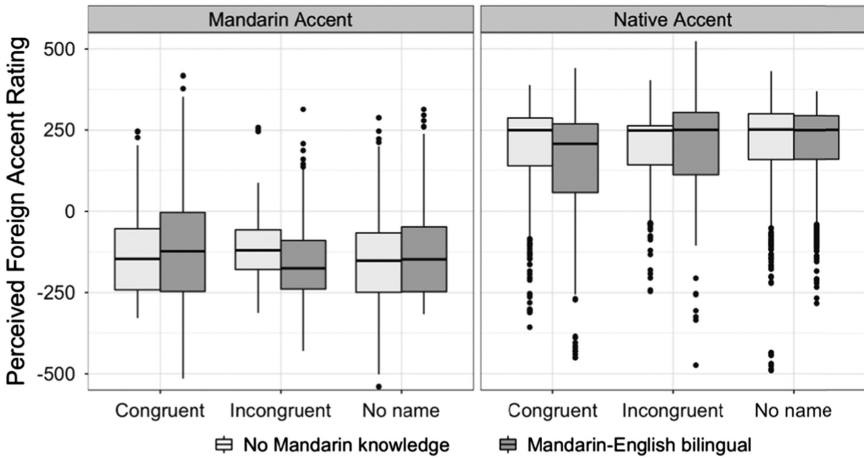


Figure 6. Boxplot of perceived foreign accentedness ratings for (a) Mandarin accents and (b) native English accents in congruent, incongruent, and no name pairings (x-axis) by listeners with no Mandarin knowledge (light gray) and Mandarin–English bilinguals (dark gray). Perceived foreign accentedness from the VAS scale is shown on the y-axis. Here, 0 denotes the visual midpoint of the scale. Negative values indicate listener judgments of strong foreign accentedness, and positive values indicate perceptions of no foreign accentedness.

the scale corresponded to strong perceived foreign accentedness and the right to no perceived foreign accent. Separate mixed effects models were run for the two accent types with normalized Visual Analogue Scale (VAS) ratings as the dependent variable. Accent/name pairing (congruent, incongruent, no name) and Mandarin experience were independent variables. There were by-listener random intercepts. Neither analysis returned any significant effects. Figure 6 provides boxplots illustrating distributions of listeners' ratings for Mandarin and native accents, respectively. This figure illustrates clear rating differences for the two accent types and the lack of clear patterns across groups.

Discussion

Experiment 2 used a within-subjects accent/name congruency manipulation where within counterbalanced blocks, listeners were presented with two Mandarin-accented voices and two native accented voices that were either associated with stereotypically congruent or incongruent names. Listeners also completed a visual analogue task where they rated the talkers' accents either with congruent names, incongruent names, or no names. In the talker recognition task, listeners were overall more accurate with the native accents, which replicates our finding from Experiment 1 with this population and voice set and confirms the basic premise of the LFE with accents. Experiment 2 also replicated our finding from Experiment 1 that the congruency of accent/name pairings matters: listeners were overall more accurate on congruent pairings than incongruent pairings,

though this effect was moderated by an interaction with listeners' experience with Mandarin. While in Experiment 1 we only found a trend toward Mandarin–English bilinguals being more accurate overall on the task, a finding that is predicted based on reported findings in the literature (e.g., Xie & Myers, 2015), Experiment 2 provides clear evidence for Mandarin–English bilinguals being more accurate on talker recognition tasks. As mentioned above, Experiment 2 showed that language experience interacted with sensitivity to the accent/name pairings. Accuracy for listeners who were Mandarin–English bilinguals was not affected by accent/name pairing. Listeners who had no Mandarin language experience, however, were more accurate on the congruent accent/name blocks, indicating socially incongruous pairing impaired listeners' abilities to recall voices.

The within-subjects accent/name congruency manipulation of this experiment allowed us to examine the kinds of errors listeners made. We approached this analysis in terms of same-accent or different-accent errors. There were more same-accent errors for the Mandarin-accented English voices, which shows that independent of the name, listeners were more likely to confuse a Mandarin-accented voice with the other Mandarin-accented voice. The less familiar accent was more confusable, again confirming a basic premise of the LFE with accents.

The error analysis also revealed an important interaction between accent/name pairings and Mandarin experience. Listeners with no Mandarin knowledge made more *same*-accent errors on congruent trials than incongruent trials. These listeners were more likely to confuse two Mandarin-accented talkers when they were assigned to stereotypically Chinese names than when those same voices were paired with stereotypically English names. Name familiarity seemingly affected the ability of listeners without Mandarin experience to recall and associate names and voices. Those with Mandarin experience had slightly lower error rates and they showed the opposite pattern: these listeners made more same accent errors when accent/name patterns were incongruent. That is, listeners with Mandarin experience were somewhat more likely to call a native English voice whose name was *Liu, Peng* and a Mandarin-accented voice whose name was *Gabriel, Luke*. Why do these two listener groups show opposing patterns with respect to the proportion of same-accent errors? We reason that Mandarin-accented voices are more challenging overall, as evidenced by the lower accuracy and longer response times to this accent. Individuals with no Mandarin language experience are likely less familiar with the stereotypically Chinese names (Heffernan, 2010), and thus find these names more challenging to associate with a voice that is also unfamiliar. This means that Mandarin-accented voices with congruent accent/name pairings will be more challenging on two fronts, accent familiarity and name familiarity, increasing same-accent error rates for the listeners with no Mandarin experience. However, Mandarin–English bilinguals who have more experience with both the accent and the names show higher same-accent error rates for the socially incongruent associations, demonstrating a preference for voices and names that match in terms of cultural association.

The accentedness rating task did not show any differences in perceived accentedness for these voices across conditions. This might be because these

voices differed greatly in terms of accentedness. Another possibility, suggested by a reviewer, is that listeners simply elected to ignore the names, as they were not deemed relevant to the task.

GENERAL DISCUSSION

These results suggest that talker recognition involves more than experience with the sound patterns of particular languages. Decades of research provide evidence that listeners show better talker recognition ability with languages and accents with which they are familiar (e.g., Goggin et al., 1991; Perrachione & Wong, 2007; Thompson, 1987; Winters et al., 2008). As in Experiment 1, the results of Experiment 2 revealed a main effect of speaker accent. All listeners more accurately recalled the native English voices than the Mandarin-accented English voices. This finding is indicative of a LFE for accents, as all listeners in our study, even those with Mandarin experience, were likely more familiar with the local accent of English than they were with Mandarin-accented English. Experiment 2 also replicated the finding from the first experiment that accent/name pairings influence how accurately listeners performed on the talker recognition task, with higher accuracy on congruent accent/name pairing trials than when the pairings were incongruent. In addition, Experiment 2 showed two effects not present in the first experiment. In Experiment 2, listeners with Mandarin experience showed better talker recognition overall, which also replicates the boost in performance for Mandarin speakers in Xie and Myers (2015). Experiment 2 also uncovered an interaction between Mandarin experience and accent/name pairing; whereas listeners with no Mandarin experience did better with congruent accent/name pairings, the Mandarin–English bilinguals performed equivalently with both types of names.

Expectations and social factors add a complicating layer in many aspects of speech processing whereby associations between subtle pronunciation patterns and regional identity (Niedzielski, 1999), a speaker's age (Drager, 2011), and apparent language background through ethnic stereotypes (Babel & Russell, 2015; Rubin, 1992) shift listeners' reported percepts. This robust sociophonetic literature illustrates that listeners analyze the speech signal differently based on social dimensions. Relatedly, in the current study we demonstrate that listeners' abilities to recognize individuals' voices is affected by socially and culturally determined accent/name pairings. Listeners were more accurate on socially congruent accent/name combinations. This effect of accent/name pairings varied according to listeners' language experiences. Listeners are less accurate at incongruent accent/name combinations, and there was some evidence that listeners with Mandarin experience made fewer errors. Our error analysis in Experiment 2, however, demonstrates that listeners with different language backgrounds made different types of errors. Listeners with Mandarin language experience were more likely to make same-accent errors on trials where names and accents were socially and culturally incongruous. Listeners without Mandarin language experience, in contrast, made more errors when the names and voices

were paired in ways that matched stereotypical social and cultural conventions. We know from the decreased accuracy with Mandarin-accented voices that these voices are more challenging to recall, and we further reason that Chinese names are harder for listeners without Mandarin language experience to associate with an unfamiliar accent. In cases where individuals have both English and Chinese names, English names are used more often by non-Chinese-speaking friends (Heffernan, 2010); this may suggest that individuals without experience speaking Mandarin are less familiar pairing voices with ethnic names. In the reverse, those with Mandarin experience may have more experience using both English names and Chinese names, so they are able to use them interchangeably while no-experience friends may only default to the English names and thus are less familiar with non-English names.

The chosen names in this study naturally differ on a number of dimensions. For example, in contrast to the multisyllabicity of some of the English names, the Chinese names were all single characters. Generally, Chinese full names consist of a one-character family surname followed by a two-character personal name (Tan, 2001), although single-character personal names are also possible and increasingly more frequent (Edwards, 2006). While there is usually a limited pool from which surnames are drawn, personal names are more numerous. Unlike in English, any character or morpheme is an acceptable Chinese given names (Edwards, 2006; Tan, 2001). English naming practices seem to follow more conventions (though consider contemporary celebrities' children's names, e.g., *Apple*, *Blue Ivy*, and transparent meaning-based names like *Grace*, *Hope*, *Rose*, etc.). Given this freer naming culture, Chinese naming practices make it difficult to select a list of "standard" Romanized Chinese first names akin to the English names selected. Moreover, as the surname is always written first, the Chinese surname is sometimes used as the English "first name" when Romanized. Given these practices, our chosen Romanized Chinese names could also be considered surnames.

In Experiment 2 we included an accentedness rating task to test whether lower talker recognition performance may be due to names affecting how the perceivable accent in the voices. Listeners' ratings of perceived accentedness of the Mandarin-accented and locally accented voices were unaffected by the name assignment. This suggests that listeners' difficulty is due to accent/name pairings and not a name-induced shift in the parsing of a voice's phonetic content.

Perrachione (in press) compares what he terms the *phonetic familiarity hypothesis* and a related *linguistic processing hypothesis*. The phonetic familiarity hypothesis hinges on passive exposure to and thus implicit familiarity with the acoustic-phonetic distributions of a language or accent, whereas the linguistic processing hypothesis requires some degree of higher level competence (e.g., word recognition and comprehension skills) on top of the lower level phonetic knowledge. While this study was not designed to adjudicate between these two related models, understanding how social knowledge might fold into a model of talker recognition likely relies on an association between phonetic and higher level linguistic representations. One way in which social information could influence talker recognition is by helping listeners anticipate the range of phonetic

variability that is likely to be germane to the task. A listener's linguistic knowledge, that is, their knowledge of words, sounds, and the combinations of sounds that compose said words in a language, provides listeners with an anticipated distribution of phonetic parameters. In this way, listeners may be able to match that anticipated distribution with the experienced distribution, thereby freeing up cognitive resources needed to associate phonetic parameters with a talker's identity. If, however, a listener does not possess such linguistic knowledge, they would need to do more bottom-up processing in order to construct a phonological category and then associate the phonetic content with a talker identity. Social expectations about a talker may function in a similar fashion. For example, if listeners are informed that they will be hearing a native speaker of the local variety of English, they are able to leverage their experience with this variety of English to anticipate the distribution of phonetic parameters, alleviating some of the cognitive load associated with "pure" bottom-up processing. Names are (nondeterministically) associated with ethnic identity and native language, and thus names will introduce their own predictions on the phonetic distributions of sounds associated with a variety of English. How broad are those predictions? If a listener has experience with individuals named *Chen* and *Peng* who do and do not speak with nonnative English accents, their anticipated distributions may be broader, more fluid, and/or less committed (distinctions that merit theorizing and empirical testing), but no less helpful. If, however, a listener's only experiences with the names *Chen* and *Peng* are of these names belonging to nonnative English speakers, their ability to usefully anticipate phonetic distributions may be attenuated due to a lack of experience with nonnative accents.

As actual experience with speaking and listening to Mandarin varies considerably within our groups, our categorization of those with and without Mandarin language experience only examines a gross and more general division. This grouping was based on their self-reported Mandarin speaking skills and lumps native speakers of Mandarin, heritage speakers of Mandarin, and second language (or third language) speakers of Mandarin into one group. Our decision to split participants based upon this self-reported metric was based on wanting to most generously model and include the student population at our university, which includes a large number of students with highly variable Mandarin experiences. The nonhomogenous nature of our listener group with Mandarin experience likely masks interesting and important differences in performance within the group. Bregman and Creel (2014), for example, found that Korean–English bilinguals who acquired English before the age of 5 were more accurate at identifying English talkers than bilinguals who learned English after this age. This suggests that early language experiences shape our strategies for listening and identifying voices in our native language. Our listeners without Mandarin experience are similarly a nonhomogenous group. These listeners come from a wide range of language backgrounds and were simply united by the lack of Mandarin speaking experience. As this study is a first look at how social factors may influence talker recognition, future research would benefit from a more fine-grained exploration of how different degrees of proficiency in a language may interact with such factors.

Our results indicate that having a Mandarin accent in an English-dominant society comes with a cost in terms of others' ability to recall one's name. This has implications for professional and personal relationships. These results also indicate that for listeners without experience with the nonnative accent, names and voices are more challenging to pair when these unfamiliar accents are associated with similarly unfamiliar names. In contrast, these nonnative voices are harder to associate with stereotypically English names for those listeners who do have experience with the nonnative accent. By no means should one extrapolate from these results any suggestion for advocacy for particular naming practices, but rather we wish to draw attention to the role of accents and names in talker recognition so that we can proactively work to attenuate any negative implications.

CONCLUSION

The results of these experiments indicate that performance in a talker recognition task can be influenced by factors extrinsic to the voices themselves, although it should be noted that these are relatively small effects, which underscores the robustness of listener experience with a language or accent sound patterns in the LFE. While a voice's name did not impact the perceived accentedness of that voice, the name assigned to a voice did influence how likely it was that the voice was remembered and later recognized.

ACKNOWLEDGMENTS

This work was supported by the Alma Mater Society's Impact Fund at UBC. Thank you to Rheanne Brownridge for help with subject running, Michelle Chan and David Kurbis for their help with stimuli selection, Qiu Ting Liu for her help with manuscript preparation, Ziya Wang for help with recording, our research participants for volunteering their time, and the speakers who lent their voices.

NOTES

1. The recording procedures for the two accent groups differed because the Mandarin speakers were being recorded for multiple purposes.
2. Variance was tested using the function `pf()` in R with the F statistic calculated as the squared standard deviation of the Mandarin-accented group divided by the squared standard deviation of the English group.
3. Five of the Mandarin-speaking participants did not provide an answer to the question "At what age did you start learning Mandarin?"; of these five, two reported Mandarin as their native language, so the actual age of Mandarin onset is likely lower.
4. Seven listeners in our no Mandarin experience group also reported some experience with a non-Mandarin tone language, but the overall experience with tone languages in this group is far lower than that of the Mandarin-English bilingual group.
5. Two of the Mandarin-speaking participants did not provide an answer to the question "At what age did you start learning Mandarin?" but did report Mandarin as their native language, so this number likely underestimates the true age of onset.

REFERENCES

- Anderson-Clark, T. N., Green, R. J., & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology, 27*, 94–99.
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of the Acoustical Society of America, 137*, 2823–2833.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991–1013.
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program]. Version 6.0.25. Retrieved February 11, 2017, from <http://www.praat.org/>
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121*, 2339–2349.
- Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*, 85–95.
- Devos, T., & Banaji, M. R. (2005). American = white? *Journal of Personality and Social Psychology, 88*, 447–466.
- Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*, 473–480.
- Drager, K. (2011). Speaker age and vowel perception. *Language and Speech, 54*, 99–121.
- Edwards, R. (2006). What's in a name? Chinese learners and the practice of adopting "English" names. *Language, Culture and Curriculum, 19*, 90–103.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110.
- Garwood, S. G., Cox, L., Kaplan, V., Wasserman, N., & Sulzer, J. L. (1980). Beauty is only "name" deep: The effect of first name on ratings of physical attraction. *Journal of Applied Social Psychology, 10*, 431–435.
- Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition, 19*, 448–458.
- Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2016). Phonological CorpusTools, Version 1.2. [Computer program]. Retrieved from <https://github.com/PhonologicalCorpusTools/CorpusTools/releases>
- Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*, 458–484.
- Heffernan, K. (2010). English name use by East Asians in Canada: Linguistic pragmatics or cultural identity? *Names, 58*, 24–36.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science, 14*, 1002–1011.
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359–384.
- Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*, 441–456.
- Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication, 10*, 265–275.
- Laham, S. M., Koval, P., & Alter, A. L. (2012). The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. *Journal of Experimental Social Psychology, 48*, 752–756.
- Leirer, V. O., Hamilton, D. L., & Carpenter, S. (1982). Common first names as cues for inferences about personality. *Personality and Social Psychology Bulletin, 8*, 712–718.

- Mehrabian, A. (2001). Characteristics attributed to individuals on the basis of their first names. *Genetic, Social, and General Psychology Monographs*, 127, 59–88.
- Munson, B., Jefferson, S. V., & McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *Journal of the Acoustical Society of America*, 119, 2427–2437.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18, 62–85.
- Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, 9, 249–260.
- Orena, A. J., Theodore, R. M., & Polka, L. (2015). Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition*, 143, 36–40.
- Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America*, 85, 913–925.
- Perrachione, T. K. (in press) Speaker recognition across languages. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception*. Oxford: Oxford University Press.
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, 333, 595.
- Perrachione, T. K., & Wong, P. C. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, 45, 1899–1910.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347–370.
- Pitt, M. A., & Szostak, C. M. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, 27, 1225–1239.
- Psychology Software Tools, Inc. [E-Prime 2.0]. 2012. Retrieved from <http://www.pstnet.com>
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33, 511–531.
- Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12, 348–351.
- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, 45, 523–538.
- Staum Casasanto, L. (2008). Does social information influence sentence processing? *Proceedings of the Cognitive Science Society*, 30, 799–804.
- Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The “other-accent” effect in voice recognition. *Journal of Cognitive Psychology*, 24, 647–653.
- Strand, E., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS Conference, Bielefeld, October, 1996* (pp. 14–26). Berlin: de Gruyter.
- Tan, P. K. W. (2001). Englishised names? *English Today*, 17, 45–53.
- Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, 1, 121–131.
- Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*, 123, 4524–4538.
- Xie, X., & Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *Journal of the Acoustical Society of America*, 137, 419–432.
- Yarmey, A. D. (1991). Descriptions of distinctive and non-distinctive voices over time. *Journal of the Forensic Science Society*, 31, 421–428.
- Yi, H. G., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *Journal of the Acoustical Society of America*, 134, EL387–EL393.

APPENDIX A

High-predictability sentences	Low-predictability sentences
People wear shoes on their feet.	We pointed at the bird.
A wristwatch is used to tell the time.	She looked at the clock.
The color of a lemon is yellow.	This is her favorite week.
The good boy is helping his mother and father.	She talked about the leaves.
A book tells a story.	Mom talked about the pie.

APPENDIX B

LANGUAGE BACKGROUND QUESTIONNAIRE

Please answer the questions below to the best of your ability. Please ask the experimenter if you have any questions or concerns.

1. Are you a native speaker of English? In this case, “native” refers to your first language. Yes / No
2. If English is not your native language, is it your dominant language? Yes / No
3. If English is not your native language, what is/are your native language(s)?
4. Regardless of whether English is your native or dominant language, what variety of English do you speak? Please specify a dialect (e.g., Newfoundland English, Southern US English, etc.) if you would like. ___ American English ___ Australian English ___ British English ___ Canadian English ___ Indian English ___ Irish English ___ Hong Kong English ___ Jamaican English ___ New Zealand English ___ Scottish English ___ Singaporean English ___ South African English ___ Other. Please, specify:
5. What gender do you identify as? _____
6. What is your racial or ethnic heritage? Check all that apply ___ First Nations ___ Asian ___ Pacific Islander ___ Black ___ White ___ Hispanic ___ South Asian ___ Other (Please specify): _____
7. What is your age? _____
8. Are you _____ right-handed or left-handed?
9. What cities or towns have you lived in? Beginning with the place where you were born, please list each town or city (and country, if appropriate) you have lived in for 6 months or more.
10. What is your proficiency in English?(1) not at all, (2) poorly, (3) fairly well, (4) fluently. Reading ___ Writing ___ Speaking ___ Listening ___
11. At what age did you start learning English? _____
12. Do you have knowledge of any languages other than English? This can include both languages you speak natively and ones you have learned in educational settings. Yes / No
13. What is your proficiency in Mandarin? If you do not have any knowledge of this language, please skip this question.(1) not at all, (2) poorly, (3) fairly well, (4) fluently. Reading ___ Writing ___ Speaking ___ Listening ___
14. At what age did you start learning Mandarin? _____

15. What other languages do you have knowledge of? Please include both languages you speak natively and ones you have learned in educational settings. When did you start learning this language? How would you rate your proficiency in reading, writing, speaking, and understanding it? (1) not at all, (2) poorly, (3) fairly well, (4) fluently.
16. Which language(s) do you most commonly speak: At home? At work? At school? With friends? With parents? With grandparents?
17. Do you have any speech or hearing disorders? If “yes,” please specify:
18. Where were your caretakers born and raised?
19. What are your caretakers’ first languages?
20. What is the highest educational degree you have earned (or are in the process of earning)?
21. What did you think the experiment was about? (optional)