Differential Neural Contributions to Native- and Foreign-Language Talker Identification

Tyler K. Perrachione, Janet B. Pierrehumbert, and Patrick C. M. Wong
Northwestern University

Humans are remarkably adept at identifying individuals by the sound of their voice, a behavior supported by the nervous system's ability to integrate information from voice and speech perception. Talker-identification abilities are significantly impaired when listeners are unfamiliar with the language being spoken. Recent behavioral studies describing the language-familiarity effect implicate functionally integrated neural systems for speech and voice perception, yet specific neuroscientific evidence demonstrating the basis for such integration has not yet been shown. Listeners in the present study learned to identify voices speaking a familiar (native) or unfamiliar (foreign) language. The talker-identification performance of neural circuitry in each cerebral hemisphere was assessed using dichotic listening. To determine the relative contribution of circuitry in each hemisphere to ecological (binaural) talker identification abilities, we compared the predictive capacity of dichotic performance on binaural performance across languages. Listeners' right-ear (left hemisphere) performance was a better predictor of binaural accuracy in their native language than a foreign one. This enhanced role of the classically language-dominant left hemisphere in listeners' native language demonstrates functionally integrated neural systems for speech and voice perception during talker identification.

Keywords: talker identification, language proficiency, laterality, hemispheric specialization, voice perception

Human beings' complex social environment has evolved hand-in-hand with two other distinct capacities: our ability to identify other people as individuals, and our ability to convey complex ideas to them via spoken language. Our person perception abilities allow us to distinguish and uniquely identify numerous other individual human beings (e.g., Bahrick, Bahrick, & Wittlinger, 1975); an ability that is realized in the auditory modality as *voice recognition* or *talker identification*. Despite the social importance of talker identification, the biological bases of this ability remain

poorly understood. A small but growing neuroscience literature has begun to identify the brain bases of voice perception abilities; however, neural studies of voice perception have generally not attempted to address a functional integration between speech- and voice-perception systems (cf. Knösche, Latner, Maess, Schauer, & Frederici, 2002; Kaganovich, Francis, & Melara, 2006). Our social auditory experiences are monopolized by listening to speech, predominately that which we can understand. In this regard, recent behavioral research has demonstrated bidirectional interactions between the cognitive abilities of speech and voice perception: Differences in linguistic proficiency affect talker identification abilities both across and within languages (Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007; Perrachione, Chiao, & Wong, 2008), and differential exposure to voices impacts speech perception abilities across talkers (e.g., Nygaard & Pisoni, 1998). Such integration across domains is the hallmark of complex neural information processing mechanisms. However, little research has examined the mechanisms by which the auditory system integrates information across cognitive domains. Talker identification abilities provide a unique opportunity to understand these integration mechanisms, as neuroscientific evidence suggests speech and voice perception abilities are predominately realized in the left and right cerebral hemispheres, respectively (e.g., Hickok & Poeppel, 2000; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). In the following study, we use dichotic listening to investigate how biological integration between the neural systems responsible for speech and voice perception may underlie the documented cognitive integration of these two abilities during talker identification. In particular, we hypothesize that compared to a foreign language, native-language talker identification tasks will

draw increasingly on linguistic processes supported by neural circuits in the left hemisphere.

The biological systems responsible for spoken language have been extensively researched. Since Carl Wernicke's seminal postmortem examinations of aphasic patients in the 1800s, we have known neural circuits in the posterior superior temporal lobe of the left hemisphere to be principally implicated in speech perception. After a century of neurolinguistic research, including the advent of functional neuroimaging, circuits in the left hemisphere remain consistently implicated as the neural basis for spoken language perception (e.g., Galuske, Schlote, Bratzke, & Singer, 2000). Contemporary work on primate evolution has furthermore pointed at the potential phylogenetic significance and underpinning of functional lateralization of the human language faculty (e.g., MacNeilage, 1990, 1998). However, little work has investigated how the neural circuits for speech perception might themselves be functionally integrated with regions or networks subserving other auditory and cognitive abilities. Recent research has begun to elucidate how, for example, music and speech processing may rely on overlapping neural circuits (Price, Thierry, & Griffiths, 2005; Wong, Skoe, Russo, Dees, & Kraus, 2007), further raising the possibility that other auditory abilities, such as voice perception, may have an intrinsic association with speech processing in the brain.

The neural systems responsible for talker identification are comparatively less well understood, with widespread investigation of this topic beginning after the discovery of voice-selective regions bilaterally in superior human temporal lobe (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Subsequent neuroimaging work has elaborated a right-hemisphere network for voice perception, predominately implicating areas in the right superior temporal sulcus (Belin, Zatorre, & Ahad, 2002; Lattner, Meyer, & Friederici, 2005; Koeda et al., 2006; Warren, Scott, Price, & Griffiths, 2006). This region has been shown to preferentially respond to human vocalizations versus other animals (Fecteau, Armony, Joanette, & Belin, 2004) as well as differentially represent familiar versus unfamiliar voices (von Kriegstein & Giraud, 2004). These contemporary neuroimaging studies agree with the results of earlier neuropsychological studies implicating the right hemisphere as the primary locus of voice perception. Van Lancker and Canter (1982) found that right-hemisphere lesions more frequently led to impairment in voice recognition than did left-hemisphere lesions. Similarly, Van Lancker and Kreiman (1987) showed that patients with right-hemisphere lesions were more impaired in the recognition of familiar voices (a condition known as *phonagnosia*), than those with left-hemisphere lesions or healthy controls. Although studies of disorders resultant from neuropsychological conditions or brain injury might reveal the system components necessary for talker identification abilities, they are inherently limited in their ability to delineate how these systems function normally in the intact brain. Moreover, studies of brain injury do not rule out a role for systems in the left-hemisphere to contribute to talker identification: Van Lancker and Kreiman (1987) further found that both left- and right-hemisphere injuries resulted in the impairment of discrimination abilities for unfamiliar voices.

Although previous functional neuroimaging studies have been successful in identifying regions that preferentially process the human voice, the specific methods they have employed have generally precluded any discovery, even incidentally, of the role of speech perception mechanisms during talker identification. Many neuroimaging studies of talker identification explicitly contrast patterns of neural activation from attending either the identity of a talker or content of the speech within the same auditory stimuli (e.g., Stevens, 2004; von Kriegstein et al., 2003). Such statistical contrasts only show differences in neural activity between speech and voice perception, and are unable to demonstrate the ways in which speech- and voice-perception systems might be functionally integrated during natural talker identification behaviors. The other method used to identify voice-selective regions contrasts vocal with nonvocal auditory stimuli (e.g., Belin et al., 2000), which confounds speech and voice into a single stimulus category. Because speech and voice information are conveyed in the same acoustic signal, it is difficult to examine the unique cognitive and perceptual effects of each, much less how they might be functionally integrated. There are, however, early indications from some neuroimaging studies hinting at how speech and voice information might be integrated. Belin et al. (2002) showed increased activation in voice-selective regions when listening to speech versus nonspeech vocal sounds. Using a functional connectivity analysis, results from von Kriegstein and Giraud (2004) indicated correlated activity between the right posterior temporal lobe and two regions in the left temporal lobe during a talker identification task. At least two electrophysiology studies have also directly confirmed that speech and voice information are integrated even in preattentive auditory processing (Kaganovich, Francis, & Melara, 2006; Knösche et al., 2002).

Besides these recent discoveries in neuroimaging, a substantial behavioral literature has consistently shown that the cognitive abilities of speech and voice perception bidirectionally integrate information between the two systems. First, there is growing behavioral evidence that in addition to structural information about a talker's voice (e.g., vocal tract length, oral cavity volume, and dynamic range of the fundamental frequency) talker identification also relies on the idiosyncratic nuances of his or her speech (phonetics). Listeners who are familiar with the language being spoken demonstrate superior talker-identification abilities to listeners unfamiliar with that language—a phenomenon known as the *language-familiarity effect* in talker identification (Goggin et al., 1991; Perrachione & Wong, 2007). Listeners with no familiarity with a foreign language appear significantly impaired in achieving native-like accuracy at identifying voices speaking that language, even after substantial training (Perrachione & Wong, 2007), which stands in contrast to other auditory abilities, such as consonant or tone identification, which can reach native-like levels after short-term training (Jamieson & Morosan, 1989; Wong & Perrachione, 2007; Wong, Perrachione, & Parrish, 2007; Song, Skoe, Wong, & Kraus, 2008). This suggests that the targets of natural talker identification are not only the structural voice cues independent of language, but also the phonetic idiosyncrasies of individual talkers. Phonetic perception is likely not a one-way street from acoustic stimulus to auditory percept; experience and expectation have also been shown to play a role in the perception of speech. Although very young infants can discriminate between nearly all human speech sounds, this ability is lost as infants become familiar with the subset of sounds relevant to their native language (Werker & Tees, 1984), and increased exposure facilitates the perception of certain phonetic distinctions by native speakers (Polka, Colantonio, & Sundara, 2002). Moreover, top-down influences on perception, such as lexical fre-

quency and the phonotactics of one's native language also add bias to phonetic perception (Davis & Johnsrude, 2007; Samuel, 1997). When listeners are identifying voices speaking a familiar language, they are able to take advantage of deviations from their stored auditory representations to help distinguish individual talkers (i.e., intertalker phonetic variability). When the language is unfamiliar, however, there are no stored representations against which to meaningfully compute variability, and therefore any phonetic cues to talker identity are either absent or unreliable. Successful integration of information from the speech-perception system has been proposed as the cognitive basis for the language-familiarity effect in talker identification (Perrachione & Wong, 2007).

A sizable speech perception literature has demonstrated that the converse is also true—variability resulting from voice can affect the perception of speech. Bradlow, Nygaard, and Pisoni (1999) showed that listeners were faster and more accurate at recognizing words spoken by familiar versus unfamiliar voices. Mullennix and Pisoni (1990) showed that as variability in an unattended feature (e.g., voices) increased, the speed of identification of an attended feature (e.g., phonetic information) decreased, suggesting significant processing effort is expended to account for talker variability during speech perception. In a compelling example of the role of stored representation in auditory perception, Johnson (1990) showed that two identical acoustic stimuli would be perceived as different vowels when listeners had prior expectations about the identity of the speaker (see also Broadbent, Ladefoged, & Lawrence, 1956). Similarly, Nygaard, Sommers, and Pisoni (1994) showed that after training to recognize the voices of individual talkers, listeners were subsequently more accurate at recognizing untrained speech tokens from those talkers. Taken together, these results support talker contingency in speech perception mechanisms, and hint at the possibility of the converse: speech-contingent mechanisms in voice recognition. Despite decades of such behavioral work showing functionally integrated systems for speech and voice perception, current models of talker identification based only on neural data fail to capture or explain this online bidirectional exchange of information (Belin, Fecteau, & Bédard, 2004; Campanella & Belin, 2007).

Here we employ dichotic listening as a measure of the underlying functional neurological configuration. The dichotic listening paradigm is predicated on the notion that stimuli presented to the left ear are primarily processed by the right hemisphere and vice versa. Dichotic listening also involves the simultaneous presentation of a distracter or masking stimulus to the unattended ear to "overwhelm" the contralateral auditory system and decrease the incidence of interhemispheric transfer effects, distinguishing this paradigm from monaural listening (Kimura, 1967). Superior behavioral performance while attending one ear is called an "ear advantage," reflecting the advantage enjoyed by the contralateral auditory system for processing a stimulus. Dichotic listening paradigms have been used extensively to assess language laterality, and researchers have consistently found right-ear (left-hemisphere) advantages for speech sounds (see Tervaniemi & Hughdal, 2003, for a review). Recent work in neuroscience has confirmed the underlying tenets of this paradigm. Dichotic listening by patients following cerebral hemispherectomy reveals virtually complete suppression of ipsilateral auditory input (de Bode, Sininger, Healy, Mathern, & Zaidel 2007). In healthy control subjects, sounds presented to either ear elicit stronger activation in the contralateral auditory network, including inferior colliculus, medial geniculate body, and auditory cortex, as measured by functional magnetic resonance imaging (fMRI) (Schönwiesner, Krumbholz, Rubsamen, Fink, & von Cramon 2007; cf. Devlin et al., 2003). Similarly, behavioral performance on speech sound discrimination as measured by dichotic listening is significantly predicted by differences in the latency of left and right N1 component of auditory-evoked potentials—an electrophysiological correlate of early cortical auditory processing (Eichele, Nordby, Rimol, & Hugdahl, 2005).

In this study, we test the hypothesis that neural circuits in the classically "language-dominant" left hemisphere play a greater role in talker identification in listeners' native language than a less familiar language. A greater role of the left hemisphere in native-language talker identification provides a neuroscientific explanation for the language-familiarity effect: Speech perception abilities supported by left hemisphere neural circuits are integrated with voice perception processes from circuits in the right hemisphere, resulting in more extensive information about talker identity and therefore superior native-language performance. If, as we predict, successful integration of information from speech perception mechanisms is the underlying factor of the language-familiarity effect, then right-ear dichotic performance should better predict overall talker identification accuracy in a native language than a nonnative one.

Method

Participants

Two groups of listeners participated in this study whose native language (L1) was either American English or Mandarin Chinese. The English L1 group consisted of 14 individuals (12 females) age 18 to 29 years ($M = 22.2$). None of the participants in the English L1 group had any familiarity with the Mandarin language. The Mandarin L1 group consisted of 13 individuals (9 females) age 18 to 31 years ($M = 23.6$). At the time of the experiment, the Mandarin L1 participants were living in the United States and had functional English-language skills, reporting between 0 and 17 years (median = 7) since their first exposure to English. All Mandarin participants reported speaking predominately Mandarin growing up. Previous studies have shown that second-language learners still exhibit a language-familiarity effect in talker identification, although they can overcome it with specific training (Perrachione & Wong, 2007). Participants were all right-handed (Oldfield, 1971) and reported no auditory or neurologic deficits. Participants gave informed written consent overseen by the Northwestern University Institutional Review Board and were compensated with a small cash payment. An additional nine participants (four Mandarin L1, five English L1) were recruited for the study but were excluded from the subsequent analysis because they performed at or below chance (~20%) on one of the six conditions described below.

Stimuli

Stimuli consisted of recordings of 10 sentences in each language condition (Mandarin or English) (Open Speech Repository, 2005), which had been used in our prior talker identification behavioral

study (Perrachione & Wong, 2007). These sentences are reproduced here in Appendix A. The English sentences were read by five male native speakers of American English (age 19-26 years, $M = 21.6$), and the Mandarin sentences were read by five male native speakers of Mandarin Chinese (age 21-26 years, $M = 22.6$). No talker read sentences in both languages, and no one who produced stimuli took part in the listening experiment. Talkers were recorded 1 to 3 years before the sample of listeners participating in this study, lessening the likelihood they came from acquainted peer groups, and no listener reported prior familiarity with any of the voices in the study. Talkers were asked to read the sentences naturally, as though they were having a conversation with a friend. Recordings were made in a sound-attenuated chamber via a SHURE SM58 microphone using a Creative USB Sound Blaster Audigy 2 NX sound card onto a Pentium IV PC. Recordings were digitally sampled at 22.05 kHz using an in-house software Wavax for Windows v2.3 and normalized for RMS amplitude to 70 dB SPL. Note that amplitude normalization likely removed overall loudness as a cue to talker identity, although individual patterns of amplitude variation are preserved. Five sentences in each language were arbitrarily designated as *training sentences*, and the remaining five as *untrained sentences*.

It bears noting that in the present experiment we intentionally used stimuli recordings of full sentences. Much of the behavioral work on the language-familiarity effect has been conducted with sentence-length stimuli or longer (Goggin et al., 1991; Thompson, 1987; Perrachione & Wong, 2007; cf. Winters, Levi, & Pisoni, 2008), whereas many of the previous behavioral studies investigating the role of voice variation on speech perception opted to use recordings of isolated words (e.g., Bradlow, Nygaard, & Pisoni, 1999). The use of sentence-length stimuli provided not only a richer phonetic environment from which to compute talker identity compared to isolated words, but also sentence-level linguistic information absent from isolated words, such as patterns of intonation, stress, and coarticulation. Additionally, work by Nygaard and Pisoni (1998) indicated listeners learned talker identity faster from sentences than single words.

Procedures

The experimental design was derived from our previous study, and has previously been shown to effectively measure talker

identification ability (Perrachione & Wong, 2007). Participants performed the task in both of two language conditions: English and Mandarin, completing the task in one language before undertaking the other. The order of conditions was counterbalanced across participants. Each condition consisted of a practice phase and a test phase, as illustrated schematically in Figure 1.

Practice Phase

During the practice phase, participants were familiarized with the voices to be recognized. Auditory stimuli were presented binaurally over headphones while participants directed visual attention to instructions on a computer monitor. Each voice read one of the five training sentences while a number designating that voice (1, 2, 3, 4, or 5) appeared on the computer monitor. After participants had heard each voice read the sentence twice, they practiced identifying the voices and received feedback. One of the five voices would read the training sentence, and the participant would indicate which voice they believed was speaking by pushing an appropriate button on a computer keyboard. The computer automatically informed the participant whether they had answered correctly, and, in the case they answered incorrectly, the computer also informed the participant of what the right answer should have been. Participants practiced identifying the voices ten times from each training sentence. Then the participants listened to the five voices reading the next sentence, and practiced identifying them from that sentence in the same way as before. This was repeated until all five voices had read all five training sentences, resulting in approximately 10 min of training altogether.

Test Phase

After practicing, participants were tested on their ability to identify the voices from the untrained sentences. Novel utterances were used to ensure participants had learned to recognize the unique features of each talker, and were not relying on more general auditory memory for the practice stimuli.

Dichotic test. Participants first identified the voices from dichotic presentation. As shown in Figure 1, they were directed to attend to one ear (the "target voice") while ignoring the other ear (the "mask voice") for blocks of 25 stimuli. The target ear was always indicated on the computer monitor during each trial. For

Practice Phase **Test Phase**

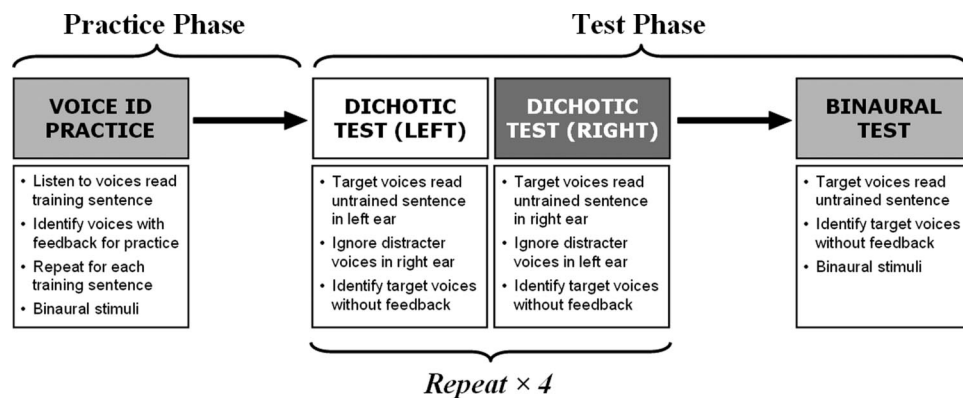| **VOICE ID PRACTICE**	**DICHOTIC TEST (LEFT)**	**DICHOTIC TEST (RIGHT)**	**BINAURAL TEST**
• Listen to voices read training sentence • Identify voices with feedback for practice • Repeat for each training sentence • Binaural stimuli	• Target voices read untrained sentence in left ear • Ignore distracter voices in right ear • Identify target voices without feedback	• Target voices read untrained sentence in right ear • Ignore distracter voices in left ear • Identify target voices without feedback	• Target voices read untrained sentence • Identify target voices without feedback • Binaural stimuli

Repeat × 4

Figure 1. Schematic representation of the phases of each language condition. The order of conditions and first ear attended was counterbalanced across participants.

each trial, the same sentence was read by these two different voices, and the target and mask voices were presented separately to each ear. For each ear, each voice served as the target an equal number of times, and each voice served as the auditory mask in the opposite ear an equal number of times. This resulted in 200 stimuli presentations in the dichotic test (5 target voices × 4 other possible mask voices × 5 sentences × 2 ears = 200 trials). The ear participants were directed to attend first was counterbalanced.

Binaural test. Participants concluded each language condition with a binaural test, which served as a measure of overall (baseline) talker identification accuracy. This test was similar to a typical listening experiment and was designed to resemble natural listening conditions, as well as be comparable to prior talker-identification studies. In our design, we opted for the binaural test to always follow the dichotic tests, rather than fully counterbalance their order (Figure 1). This design was chosen to avoid the situation where some participants had relatively more experience with the voices before beginning the dichotic condition (see, e.g., Karpicke & Roediger III, 2008, on the impact of explicit recall on learning), which may have complicated subsequent analysis and interpretation of the results. We consider performance on the binaural test to be the closest laboratory approximation of natural talker-identification ability. The same stimulus was played simultaneously to both ears while participants identified the voice. Each voice read each of the test sentences again during the binaural test, for a total of 25 trials (5 voices × 5 sentences = 25 trials). Participants did not receive feedback in either the dichotic or binaural portions of the test phase. After completing one language condition, participants were offered a short rest before repeating the task in the other condition. In total, the experiment lasted about 50 min.

Results

Participants' performance was assessed by accuracy (defined as the number of correct trials out of the total number of trials to which participants responded) and was measured separately for each ear during the dichotic test, and overall in the binaural test. To understand how neural circuitry in each hemisphere differentially contributes to ecological talker identification, we compared the predictive capacity of left-ear and right-ear accuracy on binaural performance between language conditions using correlation tests.

We also submitted participants' scores from the dichotic listening test to a repeated measures analysis of variance (ANOVA), with Ear (left vs. right) and Condition (English vs. Mandarin) as within-subject factors, and Group (English L1 vs. Mandarin L1) as a between-subjects factor. Using the ANOVA allowed us additionally to determine whether the dichotic task replicated the language-familiarity effect (Perrachione & Wong, 2007), as well as whether participants exhibited any overall ear advantages for the talker identification task.

Analysis of Differential Neural Contributions

The hypothesis we advance in this paper is that neural circuitry in the left hemisphere contributes differentially to talker-identification abilities in a familiar language compared to an unfamiliar one. This effect is independent of any overall hemispheric advantage for the task in general, a consideration we will return to below. Determining whether left-hemisphere neural circuits contribute differentially to native versus foreign language talker identification requires an explicit test between the predictive capacity of the left hemisphere on overall performance in one language condition versus its predictive capacity in the other condition. We implemented this test by first determining the correlation coefficients of performance in the left and right ears with overall performance. The statistic used in these tests was Spearman's rho (ρ) because the data were nonlinear (percents) and sometimes deviated from a normal distribution. Subsequently, we tested the relative strength of the correlation coefficients of each hemisphere across language conditions (Table 1), as well as across hemispheres within a language condition (Table 2). This analysis technique allowed us to assess the relative contribution of each hemisphere to the natural (binaural) task independent of either any neurologic left-right processing bias (i.e., the ear advantages or hemispheric asymmetries) or overall better performance in one language than the other (i.e., the language-familiarity effect), which will be considered later. (One Mandarin L1 participant was excluded from the correlation analysis in the English condition because a technical fault during the binaural test caused no data to be collected there; another was excluded from the analysis in the Mandarin condition as an outlier with performance more than 2.5 times below the standard deviation of that group's mean).

Table 1

Data Addressing the Hypothesis That Neural Circuitry in One Hemisphere Better Predicts Participants' Overall Accuracy in One Language Than the Other

Participant group	Ear	Spearman's rho (ρ)		r' (Fisher transform)		Difference (z)	p value
		English	Mandarin	English	Mandarin		
English L1	Left	0.781	0.692	1.048	0.852	0.460	>0.32
	Right	0.796	0.265	1.088	0.271	1.916	<0.03*
Mandarin L1	Left	0.606	0.760	0.703	0.996	−0.622	>0.26
	Right	0.524	0.876	0.582	1.358	−1.646	<0.05*

Note. Within-groups, between-conditions differences in each ear's predictive capacity of overall accuracy are displayed. Values in the r' (Fisher transform) column were derived from the spearman's rho (ρ) correlation coefficients, and those in difference (z) column from the r' values (see Fisher, 1921).
* Significant at p < .05 (two-tailed). For both groups, the left hemisphere (right ear) was a better predictor of overall accuracy in their native language versus the nonnative one. There was no significant difference in the predictive capacity of the right hemisphere (left ear).

Table 2

Data Addressing the Hypothesis That Neural Circuitry in One Hemisphere Is a Better Predictor of Overall Accuracy (Within a Language)

Participant group	Condition	Spearman's rho (ρ)		r' (Fisher transform)		Difference (z)	p value
		Left	Right	Left	Right		
English L1	English	0.781	0.796	1.048	1.088	−0.094	>0.46
	Mandarin	0.692	0.265	0.852	0.271	1.363	>0.08
Mandarin L1	English	0.606	0.524	0.703	0.582	0.257	>0.39
	Mandarin	0.760	0.876	0.996	1.358	−0.768	>0.22

Note. Within-groups, within-condition differences in each ear's predictive capacity for overall accuracy are displayed. Conventions are the same as Table 1.

To assess whether accuracy in either ear was a better predictor of overall accuracy, the difference (z) between each pair of normalized correlation coefficients was computed to determine whether the two correlations had the same strength. Correlation coefficients were normalized using the Fisher-z transformation, and the difference between normalized correlation coefficients was converted to a z-score (Fisher, 1921). These tests investigated whether one ear was a better predictor of overall accuracy in one language versus the other (Table 1), and whether, within a language condition, performance from one ear was a better predictor of overall accuracy than the other ear (Table 2).

English L1 participants demonstrated a significant difference in their correlation coefficients for right-ear accuracy in the English versus Mandarin conditions ($z = 1.916$, $p < .03$). For English L1 participants, accuracy in their right ear (left-hemisphere neural circuits) was a better predictor of overall accuracy when listening to their native language versus a foreign one. The same test for the left ear was not significant. Mandarin L1 participants, meanwhile, also demonstrated a significant difference in their correlation coefficients for right-ear accuracy in the English versus Mandarin conditions ($z = -1.646$, $p < .05$). For Mandarin L1 participants, accuracy in their right ear (left-hemisphere neural circuits) was also a better predictor of overall accuracy when listening to their native language than a nonnative one. Again, the same test for the left ear was not significant. Thus, for both participant groups, accuracy in their right ear (classically language-dominant left hemisphere) was a better predictor of overall accuracy in their native language than in the nonnative one. This was not the case for the left ear (right hemisphere with putative voice/indexical processing circuits). Figure 2 shows the correlations between right-ear accuracy and overall accuracy for both participant groups in each language condition. As evident from the graphs, points representing performance in either group's native language adhere much more closely to the regression model than those representing the nonnative language. There were no reliable differences between the correlation coefficients for either participant group within language condition (Table 2).

Language-Familiarity Effect

Participants' performance during the dichotic test was also submitted to a repeated measures ANOVA (see details above). Similar to prior behavioral studies of talker identification (Goggin et al., 1991; Perrachione & Wong, 2007), we found a significant Group × Condition interaction [$F(1, 25) = 46.640$, $p < 3.7 \times 10^{-7}$], indicating English L1 participants were more accurate when identifying English voices, and Mandarin L1 participants were more accurate on Mandarin voices (Figure 3). The magnitude of this effect was also similar to that of previous studies. There was no main effect of Condition, confirming neither set of voices was overall easier to identify. There was also a marginal effect of Group [$F(1, 25) = 4.072$, $p > .054$], likely owing to slightly higher performance by the Mandarin L1 participants overall.

Lateralization

The ANOVA revealed a significant Condition × Ear interaction [$F(1, 25) = 5.068$, $p < .034$], which represents a significant
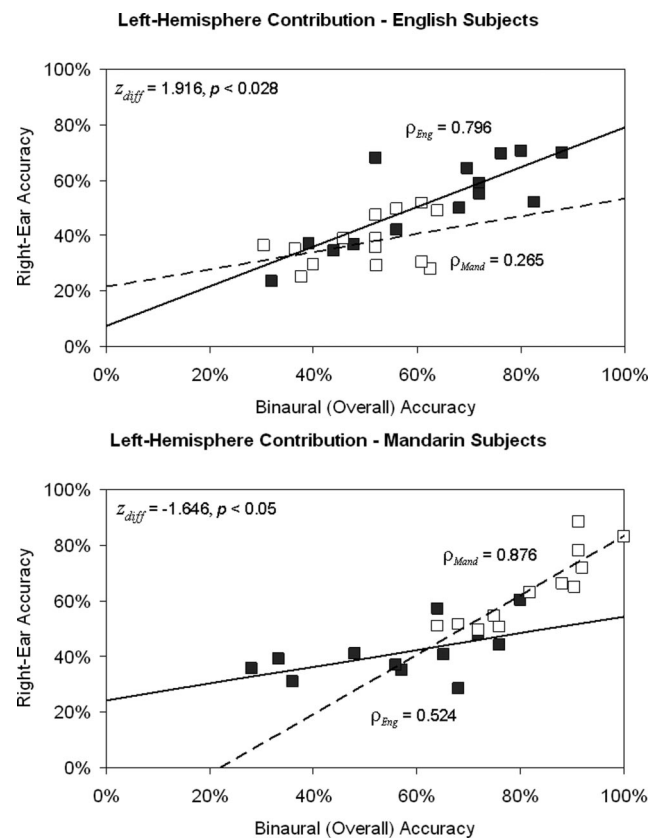

Figure 2. Predictive capacity of right-ear accuracy on overall accuracy was greater in both participant groups' native language. Filled squares (■) and a solid line represent the English condition; open squares (□) and a dashed line represent the Mandarin condition.
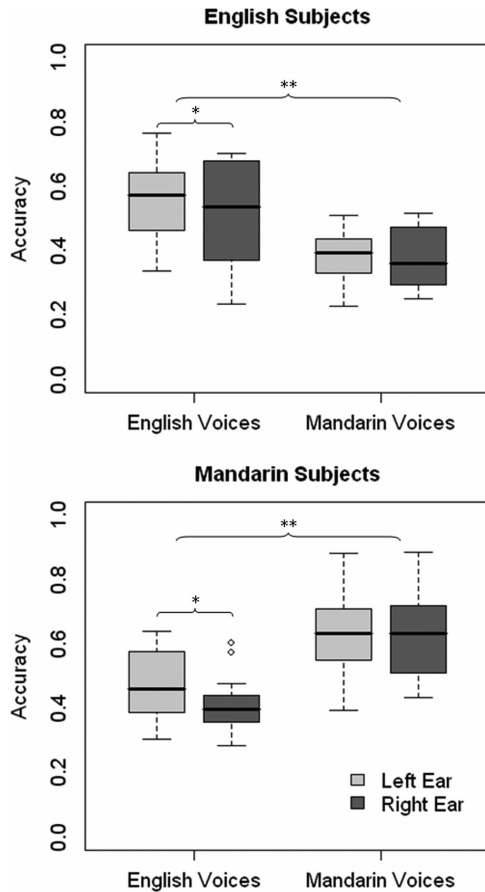
Figure 3. Boxplots of participant accuracy at voice identification. Light grey bars represent left ear (right hemisphere neural circuitry) interquartile range, dark grey bars represent the right ear (left hemisphere neural circuitry) interquartile range, high-low lines indicate maxima and minima, and open circles ($\bigcirc$) indicate data points beyond 1.5 *SD*s of the mean ($^* = p < .05$ and $^{**} = p < 4 \times 10^{-7}$).

left-ear (right cerebral hemisphere) advantage for both participant groups when identifying voices speaking English (Figure 3, left columns). There was no reliable ear-advantage for either group when listening to Mandarin. Correspondingly, there was neither a Group $\times$ Ear interaction, nor a three-way Condition $\times$ Group $\times$ Ear interaction. The main effect of Ear was also significant [$F(1, 25) = 4.413, p < .046$] and was likely driven by superior left-ear performance of both groups in the English condition. We consider possible mechanisms underlying this disparity between languages below.

Discussion

These results provide evidence for the functional integration of (classically language-dominant) left-hemisphere neural circuitry in a talker identification task and, more generally, evidence for the effects of within-modality neural integration of auditory information on human behavior. Overall talker identification accuracy was better predicted only by participants' right-ear (left-hemisphere) performance in their native language versus a non-native one. The

fact that right-ear performance differs in its predictive capacity on overall performance between languages, whereas left-ear performance does not, suggests the difference in accuracy between native- and foreign-language talker identification (the language familiarity effect) is a product of processes relying on left-hemisphere neural circuitry—most likely that of processing phonetic information related to spoken language perception.

This neuroscientific evidence for the functional integration of voice and speech in the brain during ecological talker identification is indicative of larger principles of neural organization and function. The brain has likely encountered selective pressures to develop mechanisms that maximize information about the environment, which are realized in the tendency to integrate and use available information from multiple modalities or domains whenever possible. For example, the identification of talkers by voice is enhanced when participants have both auditory and visual information available during learning (Sheffert & Olson, 2004; von Kriegstein & Giraud, 2006). Changes in auditory information can similarly affect perceptual decisions about motion, even when visual information is held constant (Ecker & Heller, 2004). In fact, numerous phenomena, both cognitive and clinical, arise when information across modalities is not integrated normally, including tinnitus (Kaltenbach, 2006), Capgras' syndrome (Hirstein & Ramachandran, 1997), and the well-known McGurk effect (McGurk & MacDonald, 1976). Here we see that even within the auditory modality, the brain integrates information from different domains (speech and voice) to optimize performance for identifying talkers, which we observe behaviorally as a language-familiarity effect.

Although carefully controlled experiments can reveal specific brain regions that are primarily dedicated to distinct functions such as speech or voice perception (e.g., von Kriegstein et al., 2003), the brain is likely to have developed to maximally integrate information from many sources to form enriched percepts. (How exactly the brain manages this task, often called the "binding problem," remains one of the most important questions in neuroscience today.) The very complexity of the human brain is defined by how these two principles, modularization and integration, have governed its development phylogenetically and ontogenetically. The language-familiarity effect, with its neural basis in the integration of speech and voice information, is one such example of how selective pressures for integrating information in general may have shaped the cognitive and perceptual systems underlying human behavior, in this case social auditory perception.

In addition to the evidence for a biological integration between speech and voice perception systems reported here, the rest of our results both replicate and extend our current understanding of the neural systems responsible for talker identification. Here we unambiguously replicated the language-familiarity effect seen in other studies of interlingual talker identification (Goggin et al., 1991; Perrachione & Wong, 2007; Thompson, 1987). It is interesting to consider the robustness of the language-familiarity effect across participant groups in light of the fact that the English L1 listeners had no familiarity of Mandarin, whereas the Mandarin L1 listeners had functional English language skills (as they were living in the United States at the time of the experiment, generally as students or employed as university researchers, or their family and friends). Both groups exhibited superior talker-identification skills in their native language, despite the familiarity of English to the Mandarin subjects. These results are consistent with our pre-

vious study (Perrachione & Wong, 2007), which used similar subject populations. In that study, Mandarin L1 subjects with some English proficiency could be trained to overcome the language-familiarity effect, whereas English L1 subjects with no Mandarin proficiency could not. The manifestation of the language-familiarity effect in these Mandarin-English bilinguals may be understood by analogy to the spatial reasoning abilities of infants. Infants may show an understanding of a concept (e.g., height, color) and the ability to use it in some situations (e.g., visual occlusion events), but are unable to make use of the same concept during another situation (e.g., a containment event) (Hespos & Baillargeon, 2007). Similarly, Mandarin-English bilinguals may be able to use their knowledge of English phonology in one situation (speech perception) without its being immediately available in a second situation (talker identification). Explicit training on English talker identification facilitates Mandarin-English bilinguals' bringing phonological information to bear in this task (Perrachione & Wong, 2007), whereas it may not be fully available without such training (present results).

Our results also suggest a general left-ear advantage for identifying voices speaking English, indicating right-hemisphere neural circuitry is optimized for this task. This result is consistent with the neuroimaging literature, which has begun to converge on regions of the right temporal lobe, especially the superior temporal sulcus, as the primary locus of the human voice-perception system (e.g., Belin et al., 2002; von Kriegstein et al., 2003). By demonstrating a left-ear advantage for a talker identification task, we show that dichotic listening paradigms are effective at eliciting a behavioral correlate of the functional lateralization of voice perception abilities. Early dichotic listening studies were largely equivocal in identifying hemispheric specialization for voice perception abilities (Doehring & Bartholomeus, 1971; Kreiman & Van Lancker, 1988; Landis, Buttet, Assal, & Graves, 1982; Tarter, 1984). This may be due in part to the challenge of working with single word stimuli, which are phonetically impoverished compared to full sentences. However, recent work, including that of Francis and Driscoll (2007), González and McLennan (2007), and the present study, has been more consistent in revealing a rightward specialization for talker-specific processing, thus providing converging evidence with results from neuroimaging and neuropsychology for the specialization of right-hemisphere neural circuitry for such tasks. The converging results from these studies may also lend empirical support to models that attribute person identification generally to neural systems situated in the right hemisphere. A growing literature exists in which right-hemisphere correlates for person identification are described in multiple domains, including Le Grand et al. (2003), Lewis et al., (2001), Neuner and Schweinberger (2000), Van Lancker and Canter (1982), and von Kriegstein et al. (2005).

The presence of an overall left-ear advantage when listening to English voices is independent of and complementary to the primary result of our study, that differential contributions of the left hemisphere between language conditions are responsible for giving rise to the language-familiarity effect. Even though the right hemisphere may be functionally optimized for talker identification tasks, our results show that differences in the availability of information provided by left hemisphere neural circuits appears to underlie the accuracy difference listeners exhibit when identifying talkers speaking a familiar versus unfamiliar language. In particular, these results suggest that native language talker identification relies on extensive integration between the information processing

mechanisms for speech and voice perception situated primarily in the left and right hemispheres, respectively. When integration between these mechanisms is impeded, such as when listening to an unfamiliar language, decreased performance reflecting the reduction in the signal's informativeness is reflected behaviorally in the language-familiarity effect.

It is additionally interesting that we did not see a similar left-ear advantage for either listener group identifying the voices speaking Mandarin. Like in English, prior research has shown a right-ear advantage for identifying Mandarin speech sounds (Ip & Hoosain, 1993; Ke, 1992), despite Mandarin being a tone language. In fact, because of the linguistic relevance of lexical tones, native Mandarin speakers exhibit a right-ear advantage for identifying lexical tones (Wang, Behne, Jongman, & Sereno, 2004; Wang, Jongman, & Sereno, 2001), a result consistent with the neuroimaging literature (Gandour et al., 2004; Wong, Parsons, Martinez, & Diehl, 2004). It is worth considering briefly why an equivalent left-ear advantage for the identification of Mandarin-speaking voices did not appear. In addition to functional differences between the left and right temporal lobes for speech and voice processing, a substantial literature has suggested these neural circuits are also sensitive to physical differences in auditory stimuli (Zatorre & Belin, 2001). The left superior temporal lobe appears to exhibit a preference for processing rapidly varying temporal features of sounds, whereas slowly varying temporal features are preferentially processed on the right (Boemio, Fromm, Braun, & Poeppel, 2005). Unlike English, which uses slowly varying intonation contours across an entire sentence, Mandarin is a tone language with different pitch contours on every syllable (Eady, 1982). It may be the case that a leftward asymmetry for processing the rapidly varying temporal envelope of Mandarin obscured the detection of any rightward asymmetry for voice perception in the current dichotic listening paradigm. Because dichotic listening is a behavioral proxy of underlying neurologic organization, it is inherently less sensitive than physiologic measures such as electroencephalography or fMRI, and thus more likely to result in Type II error (e.g., Bethmann, Tempelmann, De Bleser, Scheich, & Brechmann, 2007). For example, the magnitude of the right-ear advantage for speech can be significantly attenuated under unfavorable manipulations of attention (Mondor & Bryden, 1991). Additionally, the spatial resolution of dichotic listening paradigms is severely limited—whereas sensitive neuroimaging techniques can localize activity within millimeters, dichotic listening paradigms can only indicate left- or right-hemisphere advantages. However, such an explanation for the difference between ear advantaged for identifying English and Mandarin voices is merely speculative, and further research will be necessary to address the interaction between functional and acoustic asymmetries on dichotic listening in a hypothesis-driven manner. For example, a return to the use of isolated words as opposed to sentence-length stimuli may facilitate demonstration of such an effect because single words show substantially less differences in pitch contour dynamics between languages.

Regardless of any potential differences in processing asymmetries or the ability of dichotic listening to detect them, these results provide compelling evidence for a biological integration between neural circuitry in the left and right cerebral hemispheres for the ecological identification of talkers. Listeners who understand the language being spoken exhibit an advantage for identifying indi-

vidual talkers over listeners for whom it is unfamiliar. The language-familiarity effect for talker identification is a prominent example of how evolution has tuned the human nervous system to integrate information across domains to produce an optimal representation of the environment. The advanced social structure and complex linguistic capacity for representing and communicating the world are realized in communities where speakers and listeners predominately share the same language. As such, natural talker identification is able to draw on not only the ability to distinguish the auditory features of vocal mechanism structure, but also the ability to perceive the dynamic, phonetic idiosyncrasies of individual talkers' speech.

References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional study. *Journal of Experimental Psychology – General, 104,* 54–75.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 3,* 129–135.

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,* 309–312.

Bethmann, A., Tempelmann, C., De Bleser, R., Scheich, H., & Brechmann, A. (2007). Determining language laterality by fMRI and dichotic listening. *Brain Research, 1133,* 145–157.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8,* 389–395.

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

Broadbent, D. E., Ladefoged, P., & Lawrence, W. (1956). Vowel sounds and perceptual constancy. *Nature, 178,* 815–816.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11,* 535–543.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229,* 132–147.

de Bode, S., Sininger, Y., Healy, E. W., Mathern, G. W., & Zaidel, E. (2007). Dichotic listening after cerebral hemispherectomy: Methodological and theoretical observations. *Neuropsychologia, 45,* 2461–2466.

Devlin, J. T., Raley, J., Tunbridge, E., Lanary, K., Floyer-Lea, A., Narain, C., et al. (2003). Functional asymmetry for auditory processing in human primary auditory cortex. *Journal of Neuroscience, 23,* 11516–11522.

Doehring, D. G., & Bartholomeus, B. N. (1971). Laterality effects in voice recognition. *Neuropsychologia, 9,* 425–430.

Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech, 25,* 29–42.

Ecker, A. J., & Heller, L. M. (2005). Auditory-visual interactions in the perception of a ball's path. *Perception, 34,* 59–75.

Eichele, T., Nordby, H., Rimol, L. M., & Hugdahl, K. (2005). Asymmetry of evoked potential latency to speech sounds predicts the ear advantage in dichotic listening. *Cognitive Brain Research, 24,* 405–412.

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.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1,* 1–32.

Francis, A., & Driscoll, C. (2006). Training to use voice onset time as a cue to talker identification induces a left-ear/right-hemisphere processing advantage. *Brain and Language, 98,* 310–318.

Galuske, R. A. W., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science, 289,* 1946–1949.

Gandour, J., Tong, Y. X., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y. S., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage, 23,* 344–357.

Goggin, J., Thompson, C., Strube, G., & Simental, L. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19,* 448–458.

González, J., & McLennan, C. T. (2007). Hemispheric differences in indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 410–424.

Hespos, S. J., & Baillargeon, R. (2007). Decalage in infants' knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition, 99,* B31–B41.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92,* 67–99.

Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London Series B-Biological Sciences, 264,* 437–444.

Ip, K.-F., & Hoosain, R. (1993). Dichotic listening of Chinese and English words. *Psychologia, 56,* 140–143.

Jamieson, D. G., & Morosan, D. E. (1989). Training non-native speech contrasts in adults: Acquisition of the English /θ/–/ð/ by francophones. *Perception and Psychophysics, 40,* 205–215.

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

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research, 1114,* 161–172.

Kaltenbach, J. A. (2006). The dorsal cochlear nucleus as a participant in the auditory, attentional and emotional components of tinnitus. *Hearing Research, 216,* 224–234.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 219,* 966–968.

Ke, C.-R. (1992). Dichotic listening with Chinese and English tasks. *Journal of Psycholinguistic Research, 21,* 463–471.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3,* 163–168.

Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Frederici, A. D. (2002). Early parallel processing of auditory word and voice information. *Neuroimage, 17,* 1439–1503.

Koeda, M., Takahashi, H., Yahata, N., Asai, K., Okubo, Y., & Tanaka, H. (2007). A functional MRI study: Cerebral laterality for lexical-semantic processing and human voice perception. *American Journal of Neuroradiology, 27,* 1472–1479.

Kreiman, J., & Van Lancker, D. (1988). Hemispheric specialization for voice recognition. *Brain and Language, 34,* 246–252.

Landis, T., Buttet, J., Assal, G., & Graves, R. (1982). Dissociation of ear preference in monaural word and voice recognition. *Neuropsychologia, 20,* 501–504.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping, 24,* 11–20.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6,* 1108–1112.

Lewis, M. B., Sherwood, S., Moselhy, H., & Ellis, H. D. (2001). Autonomic responses to familiar faces without autonomic responses to fa-

miliar voices: Evidence for voice-specific Capgras delusion. *Cognitive Neuropsychiatry, 6,* 217–228.

MacNeilage, P. F. (1990). The 'Postural Origins' theory of neurobiological asymmetries in primates. In N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy, & B. Lindblom (Eds.), *Biobehavioral foundations of language development.* Hillsdale, NJ: Lawrence Erlbaum.

MacNeilage, P. F. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences, 21,* 499–546.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Mondor, T. A., & Bryden, M. P. (1991). The influence of attention on the dichotic REA. *Neuropsychologia, 29,* 1179–1190.

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

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44,* 342–366.

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

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

Oldfield, R. (1971). The assessment and analysis of handedness. *Neuropsychologia, 9,* 97–113.

Open Speech Repository. (2005). *Open Speech Repository.* Retrieved August 2005 from http://www.voiptroubleshooter.com/openspeech/index.html

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2008). *Influences of perceived racial identity on human talker identification,* 155th Meeting of the Acoustical Society of America, Paris, June 2008.

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

Polka, L., Colantonio, C., & Sundara, M. (2002). A cross-language comparison of /θ/–/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109,* 2190–2201.

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9,* 271–276.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32,* 97–127.

Schönwiesner, M., Krumbholz, K., Rubsamen, R., Fink, G. R., & von Cramon, D. Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. *Cerebral Cortex, 17,* 492–499.

Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception and Psychophysics, 66,* 352–362.

Song, J., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience, 20,* 1892–1902.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words, and tones. *Cognitive Brain Research, 18,* 162–171.

Tarter, V. (1984). Laterality differences in speaker and consonant identification in dichotic listening. *Brain and Language, 23,* 74–85.

Tervaniemi, M., & Hughdahl, K. (2003). Lateralization of auditory cortex functions. *Brain Research Bulletin, 43,* 231–246.

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

Van Lancker, D., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition, 1,* 185–195.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25,* 829–834.

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

von Kriegstein, K., & Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22,* 948–955.

von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4,* 1809–1820.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A.-L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience, 17,* 367–376.

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics, 25,* 449–466.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language, 78,* 332–348.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31,* 1389–1397.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

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.

Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insula cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience, 24,* 9153–9160.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28,* 565–585.

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping, 28,* 995–1006.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10,* 420–422.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946–953.

(Appendix follows)

Appendix A

Stimulus Sentences

English Training (Practice) Sentences

The boy was there when the sun rose.

A rod is used to catch pink salmon.

The source of the huge river is the clear spring.

Kick the ball straight and follow through.

Help the woman get back to her feet.

English Untrained (Test) Sentences

A pot of tea helps to pass the evening.

Smoky fires lack flame and heat.

The soft cushion broke the man's fall.

The salt breeze came across from the sea.

The girl at the booth sold fifty bonds.

Mandarin Training (Practice) Sentences

院子门口不远处就是一个地铁站.
yuan zi men kou bu yuan chu jiu shi yi ge di tie zhan

"There is a subway station not far from the entrance to the yard."

这是一个美丽而神奇的景象.
zhe shi yi ge mei li er shen qi de jing xiang

"This is a beautiful and magical scene."

树上长满了又大又甜的桃子.
shu shang zhang man le you da you tian de tao zi

"The treetop has grown full of big, sweet peaches."

海豚和鲸鱼的表演是很好看的节目.
hai tun he jing yu de biao yan shi hen hao kan de jie mu

"The dolphin and whale performance is a very good program to watch."

邮局门前的人行道上有一个蓝色的邮箱.
you ju men jian de ren xing dao shang you yi ge lan se de you xiang

"On the sidewalk in front of the post office there is a blue mailbox."

Mandarin Untrained (Test) Sentences

天文望远镜可以用来观察天空.
tian wen wang ya jing ke yi yong lai guan cha tian kong

"An astronomical telescope can be used to observe the sky."

她到过很多地方观光旅游.
ta dao guo hen duo di fang guang guang lu you

"She has been many places to visit."

山间的小道蜿蜒曲折.
shan jian de xiao dao wan yan qu zhe

"The mountain path winds torturously."

春天来了, 山上开满了樱花.
chun tian lai le, shan shang kai man le ying hua

"Spring has come, the mountain top has bloomed with cherry blossoms."

下雪以后, 田野里白皑皑的一片.
xia xue yi hou, tian ye li bai ai ai de yi pain

"After the snow, the field was a snow-white patch."

Stimulus sentences were taken from those made available on the Open Speech Repository (2005) database online.