INCREASED LEFT-HEMISPHERE CONTRIBUTION TO NATIVE- VERSUS FOREIGN-LANGUAGE TALKER IDENTIFICATION REVEALED BY DICHOTIC LISTENING

Tyler K. Perrachione[1,3] *and Patrick C.M. Wong*[2,4,5]

[1]Department of Linguistics, [2]Roxelyn & Richard Pepper Department of Communication Sciences and Disorders, [3]Cognitive Science Program, [4]Interdepartmental Neuroscience Program, [5]Program in Computational Biology and Bioinformatics, Northwestern University, Evanston, Illinois, USA
`pwong@northwestern.edu`

ABSTRACT

Previous studies of human listeners' ability to identify speakers by voice have revealed a reliable language-familiarity effect: Listeners are better at identifying voices when they can understand the language being spoken. It has been claimed that talker identification is facilitated in a familiar language because of functional integration between the cognitive systems underlying speech and voice perception. However, prior studies have not provided specific evidence demonstrating neural integration between these two systems.

Using dichotic listening as a means to assess the role of each hemisphere in talker identification, we show that listeners' right-, but not left-, ear (left-hemisphere) performance better predicts overall accuracy in their native than non-native language. By demonstrating functional integration of speech perception regions (classical left-hemisphere language areas) in a talker identification task, we provide evidence for a neurologic basis underlying the language-familiarity effect.

Keywords: Talker identification; Language proficiency; Laterality; Voice perception

1. INTRODUCTION

Human listeners are better at identifying voices in a familiar than unfamiliar language [1,2]. It has been proposed that an integration between speech- and voice-perception systems facilitates talker identification in one's native language [2]. When phonologically meaningful linguistic units are available for comparison against the speech signal, listeners can make use of differences between their phonemic prototypes and the phonetic nuances of an individual for identifying that talker. When the target voice is speaking an unfamiliar language, inaccurate mapping between internal phonemic representations and the acoustic-phonetic signal [3] obfuscates the relevant inter- and intra-speaker variation used in accurate talker identification.

Speech perception studies have revealed that variability due to voice affects memory for spoken words [4], the speed and accuracy of word recognition [5], and even perceived vowel quality of identical acoustic stimuli [6]. Recent behavioral work on voice perception has proposed the transfer of information between speech and voice is likely bidirectional [2]. Previous neuroimaging studies of voice perception have determined regions in the right hemisphere are primarily responsible for voice perception [7]. However, these studies often contrast activation from attending the verbal message versus attending speaker identity. This subtraction method can only identify regions unique to each perceptual system, and is unable to demonstrate how these systems might work together for talker identification tasks.

Dichotic listening tasks provide for a behavioral assessment of cerebral lateralization [8]. Stimuli presented to one ear are primarily processed by the contralateral cerebral hemisphere. For example, numerous dichotic listening studies have shown a significant right-ear advantage for perceiving speech stimuli [8]. Correspondingly, lesions to the left-hemisphere often result in linguistic deficits. Dichotic listening has only rarely been used to investigate lateralization of talker identification abilities, and the results with regards to right-hemisphere lateralization are mixed [9-11]. The present study examined the relative contributions of the left and right cerebral hemispheres to talker identification using a dichotic listening paradigm.

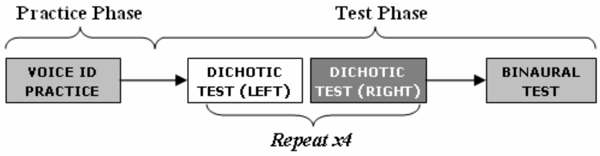
2. METHODS

2.1. Subjects

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 12 individuals (10 females) age 18 to 29 years (M = 22.2). 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 subjects were living in the United States and had functional English language skills, although all 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 [2].) Subjects were all right-handed [12] and reported no auditory or neurologic deficits. Subjects gave informed written consent overseen by the university's Institutional Review Board and received a nominal cash payment for participating.

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



2.2. Stimuli

Stimuli consisted of recordings of ten sentences in each language condition (Mandarin or English) [2]. English sentences were read by five male native speakers of American English (age 19-26, M = 21.6), and Mandarin sentences were read by five male native speakers of Mandarin Chinese (age 21-26, M = 22.6). No speaker read sentences in both languages, and no one who produced stimuli took part in the listening experiment. Sentences were digitally recorded at 22.05 kHz and normalized to 70 dB SPL RMS amplitude. Five sentences in each language were designated as "practice sentences" and the remaining five as "test sentences."

2.3. Procedure

The experiment was based on a design previously shown to effectively measure talker identification ability [2] and consisted of two language conditions: English and Mandarin. Participants completed the experiment in one language before undertaking the other, and the order of conditions was counterbalanced across subjects. Each condition consisted of a practice phase and a test phase, as illustrated schematically in Fig. 1.

During the practice phase, subjects were familiarized with the voices to be recognized. Each voice read one of the five practice sentences while a number designating that voice appeared on the computer monitor. After subjects had heard each voice read the sentence, they practiced identifying the voices with feedback indicating whether they had answered correctly or what the correct answer should have been. This was repeated until all five voices had read all five practice sentences.

After practicing, subjects were tested on their ability to identify the voices from the test sentences. Novel utterances were used to ensure subjects had learned to recognize the unique features of each voice, and were not relying on more general auditory memory for the stimuli. Subjects first identified the voices from dichotic presentation. As show in Fig. 1, they were directed to attend to the voice in one ear while ignoring the other for blocks of 25 stimuli for each ear. The target ear was always indicated on the computer monitor during each trial. Two different voices read the same sentence separately, one to each ear. For each ear, each voice served as the target an equal number of times, and each voice served as a mask for the other voices an equal number of times, resulting in 200 stimuli presentations in the dichotic test (5 voices × 4 possible distracters × 5 sentences × 2 ears = 200 trials). The ear subjects were directed to attend first was counterbalanced.

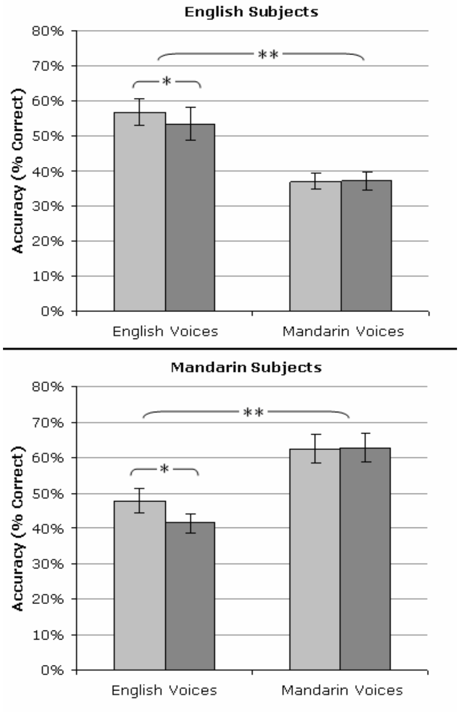
Subjects concluded each language condition with a binaural test, which served as a measure of overall talker identification accuracy. For this test, the same stimulus was played to each ear while subjects identified the voice. Each voice read each of the test sentences during the binaural test, for a total of 25 trials. Subjects did not receive feedback in either the dichotic or binaural portions of the test phase. After completing one language condition, subjects were offered a short rest before repeating the experiment in the other condition. In total, the experiment lasted about 45 minutes.

3. RESULTS

Subjects' performance was assessed by accuracy (defined as the number of correct trials out of the total number of trials to which subjects responded) and was measured separately for each ear during the dichotic test, and overall in the binaural test.

Subjects' scores were submitted to a repeated measures 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. Correlation tests were conducted on unilateral (ear) accuracy and overall (binaural) accuracy between language conditions.

Figure 2: Participant accuracy at voice identification. Light grey bars represent left ear (right hemisphere) dark grey bars represent the right ear (left hemisphere). Error bars represent standard error of the mean. ($p < 0.02$ = * and $p < 0.001$ = **).



3.1. Language-familiarity effect

Similar to prior behavioral studies of talker identification [1,2], we found a significant Group × Condition interaction [$F(1,23) = 50.024$, $p < 0.001$], indicating English L1 subjects were more accurate identifying English voices, and Mandarin L1 subjects were more accurate on Mandarin voices (Fig. 2). The magnitude of this effect was also similar to that of previous studies [2]. There was no main effect of Condition, confirming neither set of voices was overall easier to identify. There was a marginal effect of Group [$F(1,23) = 3.372$, $p = 0.079$], likely owing to slightly higher performance by the Mandarin L1 subjects.

3.2. Lateralization

The ANOVA revealed a significant Condition × Ear interaction [$F(1,23) = 6.58$, $p < 0.02$], which represents a significant left-ear (right cerebral hemisphere) advantage for both subject groups when identifying voices speaking English. There was no reliable ear advantage for either group when listening to Mandarin. The main effect of Ear was marginal [$F(1,23) = 3.611$, $p = 0.07$] and was likely driven by superior left-ear performance of both groups in the English condition.

Although no lateralization effect was observed in Mandarin, it is plausible this may be due to the more rapid temporal envelope of Mandarin speech (especially fundamental frequency – a primary cue for vocal identity). Certain models suggest left-hemisphere preference for processing stimuli changing rapidly in time [13]. Increased overall engagement of the left hemisphere in the Mandarin condition might obfuscate the extent to which voice perception is right-lateralized [7].

Table 1: Correlation coefficients (Spearman's rho) between accuracy in each ear and overall accuracy by subject group in each language condition. Asterisks indicate significant differences (p < 0.05) in predictive capacity of an ear between language conditions.

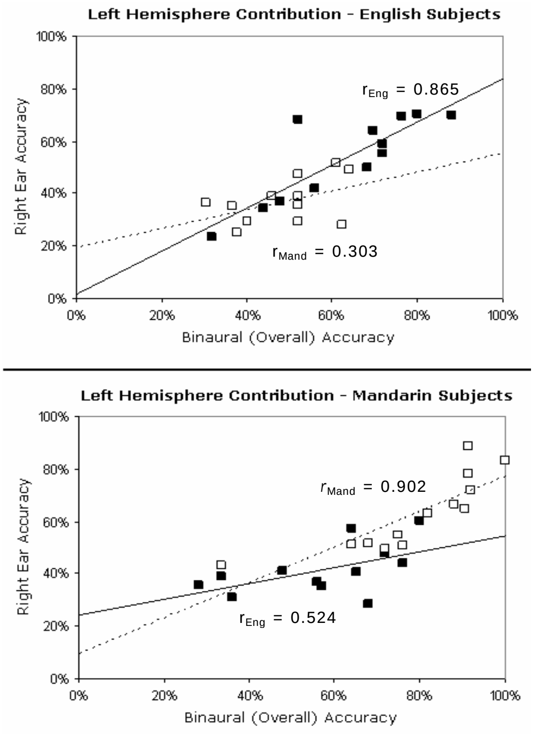
		Condition		
		Spearman's Rho		Difference (*z*)
Subject Group	Ear	English	Mandarin	
English L1	Left	0.882	0.599	1.47 (*n.s.*)
	Right	0.865	0.303	2.121 (*)
Mandarin L1	Left	0.545	0.812	-1.135 (*n.s.*)
	Right	0.524	0.902	-1.961 (*)

3.3. Hemispheric Contribution Analyses

The predictive capacity of each hemisphere on overall accuracy was assessed for each language condition by correlating subjects' accuracy between each ear with their binaural accuracy using Spearman's rho due to lack of normality. In order to determine whether the predictive capacity of either ear differed between language conditions, the difference (z) between each pair of Fisher-z transformed correlation coefficients was computed to determine whether the two correlations had the same strength. If the two correlation coefficients differed significantly in strength, then the role of that hemisphere was more closely related to overall performance in one language than the other.

As shown in Table 1, the English L1 subjects' right ear was a significantly better predictor of overall accuracy when identifying voices speaking English than Mandarin [$z = 2.121$, $p < 0.02$, 2-tailed]. Likewise, for Mandarin L1 subjects, the right ear was a significantly better predictor of overall accuracy when identifying voices speaking Mandarin than English [$z = -1.961$, $p < 0.03$, 2-tailed]. The difference in predictive capacity of the left ear between conditions was not significant for either subject group. (There was also no reliable difference between the correlation coefficients of either ear within language condition for either subject group.) Thus, accuracy from the right ear (left hemisphere) alone is a stronger predictor of overall accuracy in listeners' native language than a non-native language. This is clearly evident in Fig. 3, where the points representing performance in either group's native language adhere much more closely to the correlation line than those of the non-native language.

Figure 3: Predictive capacity of right-ear accuracy on overall accuracy was greater in both subject groups' native language. Filled squares (■) and a solid line represent the English condition; Open squares (□) and a dashed line represent the Mandarin condition.



4. DISCUSSION

These results are the first to demonstrate a specific neurologic basis underlying the language-familiarity effect in talker identification. Human listeners are relatively impaired in their ability to identify voices speaking an unfamiliar language, likely due to a functional integration between the cognitive systems responsible for speech- and voice perception [2]. The results from the present study further confirm the existence of the language-familiarity effect for talker identification. Additionally, we have shown that overall talker identification performance is significantly better predicted by accuracy on stimuli presented to the right ear when subjects listen to their native versus a non-native language, suggesting the increased contribution of the left hemisphere during talker identification specifically in one's native language.

This study is among the first to successfully use a behavioral measure to demonstrate lateralization of voice processing in English [9-11], consistent with neuroimaging studies identifying the right hemisphere as the primary locus of the voice perception system [7]. However, the focus of this study was not absolute lateralization, but the role of the left-hemisphere in talker identification. Our data revealed an increased contribution of the left hemisphere to accuracy in a native, but not foreign, language, confirming a bi-directional relationship between speech and voice perception [2,4-6].

5. REFERENCES

[1] Goggin, J., Thompson, C., Strube, G., Simental, L. 1991. The role of language familiarity in voice identification. *Memory & Cognition* 19, 448-458.

[2] Perrachione, T., Wong, P. 2007. Learning to recognize speakers of a non-native language. *Neuropsychologia* 45, 1899-1910.

[3] Best, C. 1994. The emergence of native-language phonological influence in infants. In: Goodman, J., Nusbaum, H. (eds), *The development of speech perception*. Cambridge: MIT Press, 167-224.

[4] Bradlow, A., Nygaard, L., Pisoni, D. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept. Psychophys.* 61, 206-219.

[5] Mullenix, J., Pisoni, D. 1990. Stimulus variability and processing dependencies in speech perception. *Percept. Psychophys.* 47, 379-390.

[6] Johnson, K. 1990. The role of perceived speaker identity in F0 normalization. *J. Accoust. Soc. Am.* 88, 642-654.

[7] Belin, P., Fecteau, S., Bédard, C. 2004. Thinking the voice. *Trends Cog. Sci.* 3, 129-135.

[8] Tervaniemi, M., Hughdahl, K. 2003. Lateralization of auditory cortex functions. *Brain Res. Bull.* 43, 231-246.

[9] Kreiman, J., Van Lancker, D. 1988. Hemispheric specialization for voice recognition. *Brain Lang.* 34, 246-252.

[10] Tarter, V. 1984. Laterality differences in speaker and consonant identification in dichotic listening. *Brain Lang.* 23, 74-85

[11] 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 Lang.* 98, 310-318.

[12] Oldfield, R. 1971. The assessment and analysis of handedness. *Neuropsychologia.* 9, 97-113.

[13] Zatorre, R., Belin, P., Penhume, V. 2002. Structure and function of auditory cortex. *Trends Cog. Sci.* 6, 37-46.

This work is supported by the National Institutes of Health (U.S.A.) grants HD051827 & DC007468 awarded to P.W.