Human Voice Recognition Depends on Language Ability

Tyler K. Perrachione,[1]* Stephanie N. Del Tufo,[1] John D. E. Gabrieli[1,2]*

The ability to recognize individual conspecifics from their communicative vocalizations is an adaptive trait evinced widely among social and territorial animals, including humans. Studies of human voice recognition compare this ability to nonverbal processes, such as human perception of faces or nonhuman animals' perception of vocalizations (*1*). However, the human voice is also the principal medium for the human capacity of language, as conveyed through speech. Human listeners are more accurate at identifying voices when they can understand the language being spoken (*2*), an advantage thought to depend on listeners' knowledge of phonology—the rules governing sound structure in their language. Leading theories of dyslexia propose that impoverished phonological processing often underlies impaired reading ability in this disorder (*3*, *4*). We therefore hypothesized that, if voice recognition by human listeners relies on linguistic (phonological) representations, listeners with dyslexia would be impaired compared with control participants when identifying voices speaking their native language (because of impaired phonological processing) but unimpaired in voice recognition for an unfamiliar, foreign language (where both individuals with and without dyslexia lack relevant language-specific phonological representations).

We assessed participants with and without dyslexia for their ability to learn to recognize voices speaking either the listener's native language (English) or an unfamiliar, foreign language (Mandarin Chinese). In each language, participants learned to associate five talkers' voices with unique cartoon avatars and were subsequently tested on their ability to correctly identify those voices. The participants' task was to indicate who of the five talkers spoke in each trial [five-alternative forced choice; chance = 20% accuracy (*5*)]. Despite using the same vocabulary, all speakers of a language differ in their pronunciations of words (*6*), and listeners can use their phonological abilities to perceive these differences as part of a speaker's vocal identity. A repeated-measures analysis of variance revealed that, compared with controls, dyslexic participants were significantly impaired at recognizing the voices speaking English but unimpaired for those speaking Chinese (group × condition interaction, $P < 0.0006$) (Fig. 1).

English-speaking listeners with normal reading ability were significantly more accurate identifying voices speaking English than Chinese (paired t test, $P < 0.0005$), performing on average 42% better in their native language (*7*). English-speaking listeners with dyslexia were no better able to identify English-speaking voices than Chinese-speaking ones (paired t test, $P = 0.65$), with an average performance gain of only 2% in their native language. Correspondingly, dyslexic listeners were significantly impaired compared with controls in their ability to recognize English-speaking voices (independent-sample t test, $P < 0.0021$). Dyslexic listeners were as accurate as controls when identifying the Chinese-speaking voices (independent-sample t test, $P = 0.83$), demonstrating that their voice-recognition deficit was not due to generalized auditory or memory impairments. Moreover, for the dyslexic participants, greater impairments on clinical assessments of phonological processing were correlated with worse accuracy for identifying English-speaking voices (both Pearson's $r > 0.6$, $P < 0.015$). Although the diagnostic criterion for dyslexia is impairment in developing typical reading abilities, these data show that reading difficulties are accompanied by impaired voice recognition. This inability to learn speaker-specific representations of phonetic consistency may reflect a weakness in language learning that contributes to impoverished long-term phonological representations in dyslexia.

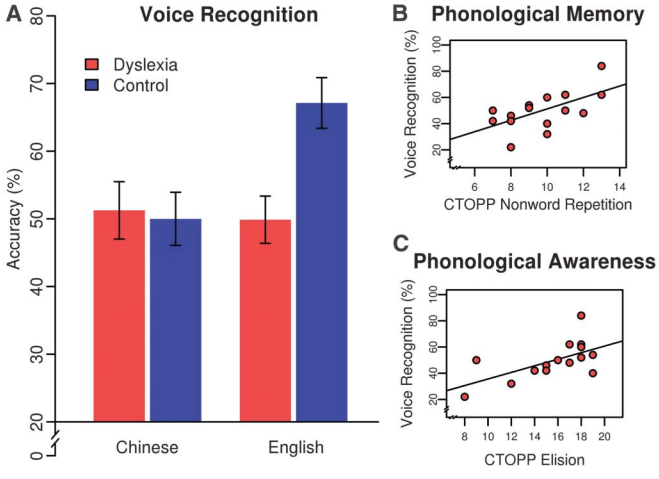

Fig. 1. (**A**) Mean voice-recognition performance of dyslexic and control listeners (error bars indicate SEM). All individuals scored above chance (20%), shown as baseline. (**B** and **C**) Relationships between clinical measures of language (phonological) ability in dyslexia and voice-recognition ability. CTOPP, Comprehensive Test of Phonological Processing.

For humans, the ability to recognize one another by voice relies on the ability to compute the differences between the incidental phonetics of a specific vocalization and the abstract phonological representations of the words that vocalization contains. When the abstract linguistic representations of words are unavailable (because the stimulus is unfamiliar, as in foreign-language speech) or impoverished (because native-language phonological representations are compromised, as in dyslexia), the human capacity for voice recognition is significantly impaired. This reliance on our faculty for language distinguishes human voice recognition from the recognition of conspecific vocalizations by other nonhuman animals.

References and Notes

1. P. Belin, S. Fecteau, C. Bédard, *Trends Cogn. Sci.* **8**, 129 (2004).
2. T. K. Perrachione, P. C. M. Wong, *Neuropsychologia* **45**, 1899 (2007).
3. L. Bradley, P. E. Bryant, *Nature* **301**, 419 (1983).
4. J. D. E. Gabrieli, *Science* **325**, 280 (2009).
5. Materials and methods are available as supporting material on *Science* Online.
6. J. Hillenbrand, L. A. Getty, M. J. Clark, K. Wheeler, *J. Acoust. Soc. Am.* **97**, 3099 (1995).
7. Native Chinese-speaking controls exhibit the opposite pattern, recognizing Chinese-speaking voices more accurately than English-speaking ones (*2*), revealing the critical factor to be listeners' language familiarity, not properties inherent to the voice stimuli or languages themselves.

Acknowledgments: We thank J. A. Christodoulou, E. S. Norton, B. Levy, C. Cardenas-Iniguez, J. Lymberis, P. Saxler, P. C. M. Wong, C. I. Moore, and S. Shattuck-Hufnagel. This work was supported by the Ellison Medical Foundation and NIH grant UL1RR025758. T.K.P. is supported by an NSF Graduate Research Fellowship.

Supporting Online Material

www.sciencemag.org/cgi/content/full/333/6042/595/DC1
Materials and Methods
Fig. S1
Table S1
References (*8*–*16*)

21 April 2011; accepted 16 June 2011
10.1126/science.1207327

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]McGovern Institute for Brain Research and Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: tkp@mit.edu (T.K.P.); gabrieli@mit.edu (J.D.E.G.)

Supporting Online Material for

Human Voice Recognition Depends on Language Ability

Tyler K. Perrachione,* Stephanie N. Del Tufo, John D. E. Gabrieli*

*To whom correspondence should be addressed. E-mail: tkp@mit.edu (T.P.); gabrieli@mit.edu (J.D.E.G.)

Published 29 July 2011, *Science* **333**, 595 (2011)
DOI: 10.1126/science.1207327

This PDF file includes:

Materials & Methods

Participants

Native English-speaking controls (N = 16; 9 female) 18-30 years of age (M = 21.2; SD = 2.98) with a self-reported history free from neurological, psychiatric, speech, language, or reading impairments were matched with individuals with dyslexia (N = 16; 8 female) between 16-38 years of age (M = 24; SD = 6.8). Inclusionary criteria for dyslexia consisted of a prior clinical diagnosis or lifelong history of reading disability and scoring below the 16th percentile (one standard deviation below the age-normed mean) on any two subtests from the following standard clinical reading and language assessments: *Woodcock Reading Mastery Test-Revised* (WRMT-R/NU) (7), *Test of Word Reading Efficiency* (TOWRE) (8), and *Comprehensive Test of Phonological Processing* (CTOPP) (9). Groups were matched based on cognitive performance ("Matrices" and "Block Design" from the *Wechsler Abbreviated Scale of Intelligence*, WASI; (10)), working memory (*Wechsler Adult Intelligence Scale* WAIS-IV; (11)), age, and education. All participants were right-handed based on questionnaire responses (adapted from the Edinburgh Handedness Inventory; (12)). All participants indicated no prior experience with Mandarin Chinese. Informed written consent, approved by the MIT Committee on the Use of Humans as Experimental Subjects, was obtained from all participants prior to participation.

Stimuli

Two sets of ten sentences designed for acoustic assessment were recorded for this experiment: one spoken in English (13), the other in Mandarin (14). The English

sentences were read by five male native speakers of American English (aged 19-26 years, *M* = 21.6). The Mandarin sentences were read by five male native speakers of Mandarin Chinese (aged 21-26 years, *M* = 22.6). No talker read sentences in both languages, and none of the individuals recorded as talkers participated in the listening experiment. Recordings were made in a sound-attenuated chamber via a SHURE SM58 microphone using a Creative USB Sound Blaster Audigy 2 NX sound card, sampled at 22.05 kHz and normalized for RMS amplitude to 70 dB SPL. Recordings of sentences were 1.46sec to 4.09sec in duration (*M* = 2.43, *SD* = 0.54). In each language, five sentences were used during the familiarization and practice phases, and all ten were used during the final voice recognition test. These stimuli have been used in prior experiments of voice recognition by native speakers of English and Chinese (2,15).

Procedure

Participants learned to identify five talkers in each of two language conditions (English and Mandarin) from the sound of their voice. Each talker was associated with a distinct cartoon avatar (Fig. S1). Training and testing on voice recognition were completed in each language condition separately, and the order was counterbalanced across listeners. During an initial familiarization phase, participants heard each of the voices in succession while the corresponding avatars were displayed on a computer screen. Participants then actively practiced identifying the talkers with corrective feedback: The five avatars appeared on the screen while a recording from one talker was played, and participants selected the avatar matching the voice they heard. If participants selected incorrectly, the computer indicated the correct response. During the task, all

instructions were presented both as text on the screen and as auditory prompts recorded by an additional female talker. The familiarization and active practice phases were repeated over five training sentences, and each sentence was practiced ten times. Following training, participants undertook a 50-item talker identification test, in which they identified the voices without feedback. Participants completed the self-paced experiment in a quiet room. Stimuli were presented binaurally at a comfortable level over Sennheiser HD-250 linear II circumaural headphones using an Edirol UA-25EX sound card.

References and Notes

1. P. Belin, S. Fecteau, C. Bédard, Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci.* **8**, 129 (2004). doi:10.1016/j.tics.2004.01.008 Medline

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

3. L. Bradley, P. E. Bryant, Categorizing sounds and learning to read—a causal connection. *Nature* **301**, 419 (1983). doi:10.1038/301419a0

4. J. D. E. Gabrieli, Dyslexia: A new synergy between education and cognitive neuroscience. *Science* **325**, 280 (2009). doi:10.1126/science.1171999 Medline

5. Materials and methods are available as supporting material on *Science* Online.

6. J. Hillenbrand, L. A. Getty, M. J. Clark, K. Wheeler, Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**, 3099 (1995). doi:10.1121/1.411872 Medline

7. Native Chinese-speaking controls exhibit the opposite pattern, recognizing Chinese-speaking voices more accurately than English-speaking ones (*2*), revealing the critical factor to be listeners' language familiarity, not properties inherent to the voice stimuli or languages themselves.

8. R. W. Woodcock, *Woodcock Reading Mastery Tests – Revised/Normative Update (WRMT-R/NU)* (American Guidance Service, Circle Pines, MN, 1998).

9. J. K. Torgesen, R. Wagner, C. Rashotte, *Test of Word Reading Efficiency (TOWRE)* (Pro-Ed, Austin, TX, 1999).

10. R. Wagner, J. K. Torgesen, C. Rashotte, *Comprehensive Test of Phonological Processing (CTOPP)* (Pro-Ed, Austin, TX, 1999).

11. D. Wechsler, *Wechsler Abbreviated Scale of Intelligence* (Psychological Corporation, San Antonio, TX, ed. 3, 1999).

12. D. Wechsler, *Wechsler Memory Scale* (Pearson, San Antonio, TX, ed. 4, 2008).

13. R. C. Oldfield, The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97 (1971). doi:10.1016/0028-3932(71)90067-4 Medline

14. Institute of Electrical and Electronics Engineers, IEEE recommended practices for speech quality measurements. *IEEE Trans. Audio Electroacoust.* **17**, 225 (1969). doi:10.1109/TAU.1969.1162058

15. Open Speech Repository, www.voiptroubleshooter.com/open_speech/index.html.

16. T. K. Perrachione, J. B. Pierrehumbert, P. C. M. Wong, Differential neural contributions to native- and foreign-language talker identification. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1950 (2009). doi:10.1037/a0015869 Medline

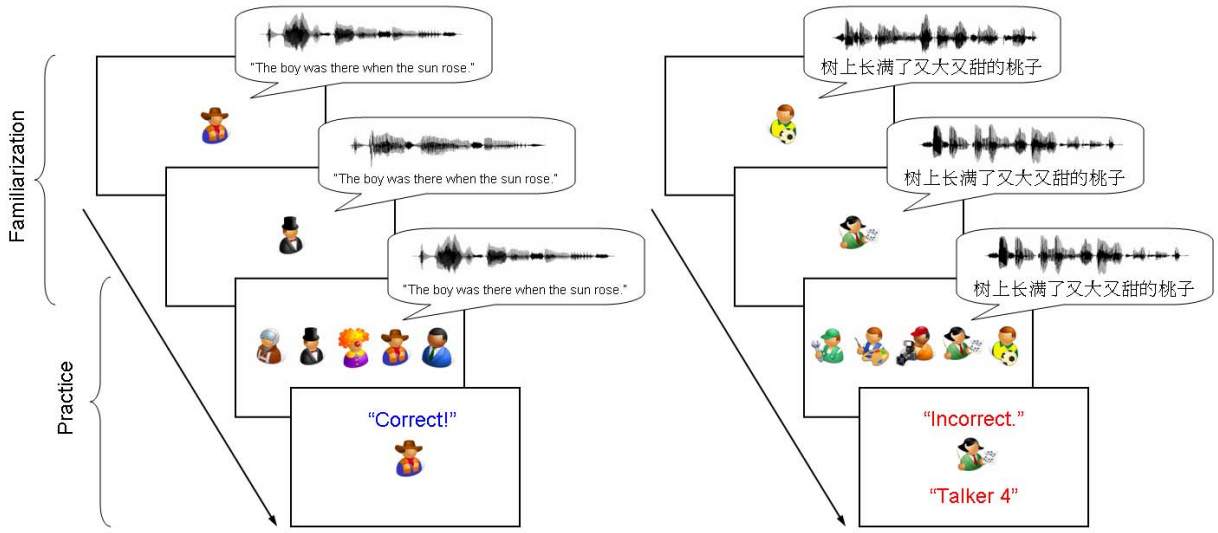

Figure S1: Graphical depiction of training paradigm. Listeners learned to recognize 5 English and 5 Chinese talkers from the sound of their voice. Talkers were paired with distinct icons, and acoustic waveforms illustrate variability in talker characteristics. Participants were familiarized with individual talkers and practiced recognizing them with feedback. Talker identification accuracy in each condition was assessed with a post-training test.

		Control		Dyslexia		
Test	Subtest	Raw Score	Standard Score	Raw Score	Standard Score	Cohen's *d*
WASI						
	Block Design	60.6 ± 7.4	61.4 ± 5.4	53.9 ± 13.4	57.1 ± 8.1	0.644
	Matrix Reasoning	29.8 ± 2.3	58.6 ± 4.7	29.8 ± 3.8	58.4 ± 67.5	0.022
	Performance IQ	119.9 ± 8.5	116.8 ± 8.8	115.5 ± 11.0	112.2 ± 11.0	0.487
CTOPP						
	Elision	18.9 ± 1.5	10.9 ± 1.5	15.7 ± 3.4	8.3 ± 22.7	1.401
	Blending Words	18.5 ± 1.4	12.5 ± 1.4	14.1 ± 3.5	8.9 ± 2.5	1.833
	Non-word Repetition	15.7 ± 2.1	11.8 ± 1.9	9.7 ± 1.9	6.8 ± 1.3	3.134
	Rapid Digit Naming	22.9 ± 5.1	10.4 ± 2.6	31.2 ± 8.6	6.9 ± 2.8	1.315
	Rapid Letter Naming	22.8 ± 4.5	11.1 ± 2.7	35.4 ± 8.9	5.5 ± 2.9	2.072
	Rapid Object Naming	39.4 ± 5.6	10.6 ± 3.1	51.7 ± 8.5	6.3 ± 2.1	1.721
WRMT-R/NU						
	Word ID	101.7 ± 2.8	112.4 ± 8.1	91.3 ± 6.9	94.2 ± 7.8	2.368
	Word Attack	42.4 ± 1.7	121.3 ± 13.3	31.6 ± 4.4	92.0 ± 8.0	2.762
TOWRE						
	Sight Word Reading	99.0 ± 8.3	106.1 ± 11.3	79.9 ± 15.4	85.3 ± 13.0	1.765
	Decoding	54.7 ± 6.1	104.8 ± 11.4	35.4 ± 10.7	76.7 ± 18.3	1.899
	Total	200.0 ± 40.9	106.5 ± 11.6	166.2 ± 21.9	75.4 ± 19.8	1.975
WAIS-IV						
	Digit Span Total	8.6 ± 1.5	9.6 ± 1.9	7.9 ± 1.9	8.8 ± 2.5	0.403
Age (years)		21.3 ± 2.7		23.9 ± 6.8		0.536
Education (years)		15.3 ± 1.5		15.1 ± 2.4		0.286

Table S1: Cognitive and behavioral assessment profile of participants. Dyslexic and control participants were matched on performance IQ, working memory, age, and education, but differed on measures of reading ability and phonological processing. Values indicate mean ± standard deviation. Abbreviations: WASI: *Wechsler Abbreviated Scale of Intelligence, 3rd Ed.* (10); CTOPP: *Comprehensive Test of Phonological Processing* (8); WRMT-R/NU: *Woodcock Reading Mastery Test-Revised* (6), TOWRE: *Test of Word Reading Efficiency* (8); WAIS-IV: *Wechsler Adult Intelligence Scale* (9). Cohen's *d* shows the effect size of the group difference in standard scores.