PHONETIC VARIABILITY IN SPEECH PERCEPTION AND THE PHONOLOGICAL DEFICIT IN DYSLEXIA

Tyler K. Perrachione[a], *Stephanie N. Del Tufo*[a], *Satrajit S. Ghosh*[b] *& John D.E. Gabrieli*[a]

[a]Department of Brain and Cognitive Sciences; [b]Research Laboratory of Electronics, Massachusetts Institute of Technology, USA
tkp@mit.edu

ABSTRACT

Accounting for phonetic variability across talkers is a core challenge in speech perception. Cognitive models of word recognition address variability by employing either episodic or talker-normalization based approaches. In developmental dyslexia, a "phonological deficit" is thought to impair the acquisition of typical reading ability; however, no connection has yet been made between such a phonological deficit and speech perception behavior, which appears intact in dyslexia. We demonstrate differences between dyslexic listeners and controls in two tasks involving processing phonetic variability: First, dyslexics exhibited impaired talker identification abilities in a familiar language but not a foreign one. Second, the brains of individuals with dyslexia exhibited reduced hemodynamic adaptation – a neural index of talker normalization – compared to typical readers.

Together, these results suggest speech perception in dyslexia relies primarily on episodic processes, and that the inability to fully utilize talker-normalization processes may impair the formation of the abstract representations of speech sounds necessary for robust sound-to-letter mapping during reading development.

Keywords: talker normalization, dyslexia, speech perception, talker identification, adaptation

1. INTRODUCTION

Developmental dyslexia is a neurological condition that impairs the development of reading skills despite otherwise normal intelligence. The most widely acknowledged source of this impairment comes from a so-called "phonological deficit", in which the representation of, or access to, abstract speech sounds is presumed to be somehow deficient [6]. However, no hypothesis has presented a convincing account of the source of such a deficit. This is especially problematic given the equivocal evidence for speech-perception deficits in dyslexia [2], which would seem to follow necessarily from any disorder of phonological representation or access.

Models of speech perception differ in the extent to which they rely on abstract phonemic units for lexical access. Exemplar-based models [1] rely on massive episodic memory networks to compare incoming stimuli to stored traces of previous experiences, while talker-normalization models [4] entail the dynamic construction of "transfer functions" between the phonetic nuances of an individual talker's speech and listeners' long-term abstract representations of speech sounds. There is substantial evidence to suggest both mechanisms play a role during typical language processing.

Because underlying abstract representations are explicitly specified in talker-normalization models, we hypothesized that, if there is an underlying phonological deficit, individuals with dyslexia should be less able to draw on talker-normalization processes during speech perception behaviors. Aspects of speech perception abilities in dyslexia may appear unimpaired due primarily to their reliance on episodic processes, whereas typical reading development may require the underlying abstract representations associated with talker normalization.

2. EXPERIMENT 1: PHONOLOGICAL PROCESSES IN VOICE RECOGNITION

2.1. Methods

Previous work has shown that listeners are more accurate at identifying voices speaking a familiar language than a foreign one [5], likely because they detect consistent deviations between a talker's phonetics and their abstract representations when phonemic units are known. We hypothesized that, due to impoverished abstract representations, listeners with dyslexia would not exhibit this "language familiarity effect" for enhanced native-language talker identification.

2.1.1. Participants

A group of native English-speaking young-adult controls (N=16) with no known neurological or psychiatric impairments and a self-reported history free from speech or language problems, as well as a second group with a prior dyslexia diagnosis or history of reading difficulties (N=16) were recruited for this study. Participants were assessed for their performance on a battery of standard reading and phonological measures (Table 1), to confirm their status as typical or impaired readers.

Table 1: Behavioral assessment standard scores in Experiment 1; Cohen's *d* gives the effect size of the group difference.

Assessment	Control	Dyslexia	*d*
WASI			
Performance IQ	117 ± 8	112 ± 11	0.5
CTOPP			
Elision	11 ± 2	8 ± 2	1.4
Blending	13 ± 1	9 ± 2	1.8
Nonword Repetition	12 ± 2	7 ± 1	3.1
WRMT - Revised			
Word ID	112 ± 8	94 ± 8	2.4
Word Attack	121 ± 13	92 ± 8	2.8
TOWRE			
Sight Word Efficiency	106 ± 11	85 ± 13	1.8
Phonemic Decoding	105 ± 11	77 ± 18	1.9
WAIS - IV			
Digit Span (Total)	10 ± 2	9 ± 3	0.4

2.1.2. Stimuli

Two sets of ten sentences designed for acoustic assessment were recorded for this experiment: one spoken in English, the other in Mandarin. The English sentences were read by five young adult male native speakers of American English. The Mandarin sentences were read by five young adult male native speakers of Mandarin Chinese. No talker read sentences in both languages, and none of the talkers participated in the listening experiment. Recordings were made in a sound-attenuated chamber via a SHURE SM58 microphone and Creative USB Sound Blaster Audigy 2 NX sound card, sampled at 22.05 kHz and normalized for RMS amplitude to 70 dB SPL. Sentence recordings were 2.43 ± 0.54s in duration. In each language, five sentences were used for familiarization and practice, and all ten were used during the final voice recognition test.

2.1.3. Procedure

Participants learned to identify five talkers in each of the two language conditions from the sound of their voice. Each talker was associated with a distinct cartoon avatar. Training and testing on voice recognition were completed in each language condition separately, with order counterbalanced across listeners. In 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.

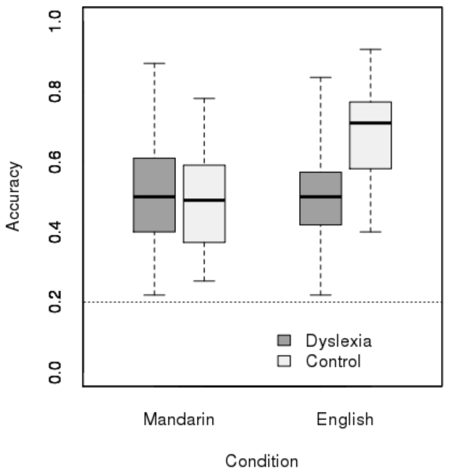
2.2. Results & discussion

Participants' voice-recognition accuracy was analyzed with a 2×2 repeated-measures ANOVA for effects of Condition (English vs. Mandarin) and Group (control vs. dyslexia). Dyslexic participants exhibited significantly impaired voice recognition abilities compared to controls in English, but not Mandarin (Fig. 1); [Group × Condition interaction; $F_{1,30}$=14.82, p<0.0006]. Pairwise comparisons (all two-tailed, and paired- or independent-sample as appropriate) revealed that: **(*1*)** controls identified voices more accurately in English than Mandarin [t_{15}=4.52, p<0.0005]; **(*2*)** controls identified voices speaking English more accurately than the dyslexic participants did [t_{30}=3.37, p<0.0021]; **(*3*)** dyslexic participants' performance did not differ between languages [t_{15}=0.47, p=0.65.]; and **(*4*)** the two groups did not differ in their ability to identify Mandarin voices [t_{30}=0.22, p=0.83]. Performance was above chance (20%) for all participants.

For dyslexic participants, standard clinical measures of phonological processing ability correlated positively with their ability to recognize voices speaking English [phonological encoding: r=0.60, p<0.015; phonological awareness: r=0.61, p<0.012], but not Mandarin [both r<0.33]. No such relationship was seen in controls.

In native-language talker identification, the transfer functions produced by talker normalization to facilitate speech perception can be "inverted" by listeners to make phonetic consistency a cue to talker identity. Given their significant impairment relative to controls for native-, but not foreign-language talker identification, individuals with dyslexia appear unable to use talkers' phonetic consistencies to compute their identity – a process that relies on stored abstract representations [5].

Figure 1: Compared to controls, individuals with dyslexia are impaired at identifying voices in their native language, but not a foreign one.



3. EXPERIMENT 2: NEURAL SIGNATURE OF TALKER NORMALIZATION

3.1. Methods

Talker-normalization processes in speech are a specific case of general-purpose neural adaptation mechanisms that reduce physiologic cost and make processing more efficient when information about the environment is consistent and predictable. We hypothesized that, compared to typical readers, individuals with dyslexia would exhibit reduced neurophysiologic adaptation to a repeated talker during a speech perception task (as measured by functional magnetic resonance imaging, fMRI).

3.1.1. Participants

A group of native English-speaking young-adult controls (N=17) with no know neurological or psychiatric impairments and a self-reported history free from speech or language problems, as well as a second group with a prior dyslexia diagnosis or history of reading difficulties (N=18) were recruited for this study. Participants were assessed for their performance on a battery of standard reading and phonological measures (Table 2), to confirm their status as typical or impaired readers.

Table 2: Behavioral assessment standard scores in Experiment 2; Cohen's *d* gives the effect size of the group difference.

Assessment	Control	Dyslexia	*d*
WASI			
Performance IQ	121 ± 6	109 ± 13	1.2
CTOPP			
Elision	12 ± 1	8 ± 2	2.6
Blending	12 ± 1	10 ± 3	1.0
Nonword Repetition	12 ± 3	7 ± 1	2.4
WRMT - Revised			
Word ID	113 ± 12	91 ± 7	2.4
Word Attack	112 ± 13	92 ± 7	2.0
TOWRE			
Sight Word Efficiency	108 ± 8	84 ± 7	3.2
Phonemic Decoding	105 ± 10	80 ± 7	2.9
WAIS - IV			
Digit Span (Total)	14 ± 4	9 ± 4	1.3

3.1.2. Stimuli

Five female native-English speakers were recorded reading 288 monosyllabic nouns in isolation via a SHURE SM58 microphone using an Edirol UA-25EX sound card, sampled at 44.1 kHz and normalized for RMS amplitude to 70 dB SPL. Recordings of words were 0.53 ± 0.15s in duration. Of these words, 36 were selected as targets, to be depicted by black and white line drawings during the in-scanner behavioral task.

3.1.3. Behavioral task

One of the target images appeared on the screen for 11s while eight audio recordings were played in succession. Participants' task was to press a button indicating when the word they heard matched the picture they saw. This task was performed alternately under two conditions: **(*1*)** low-variability (LV), in which the auditory stimuli were produced by a single, consistent talker; and **(*2*)** high-variability (HV), in which the auditory stimuli were produced by four different talkers.

Neural processes related to talker normalization were predicted in the LV condition, because indexical features were consistent and predictable. The HV condition did not permit talker normalization (or its neural signature, adaptation) because four voices were mixed with no trial-by-trial predictability in indexical features.

Blocks of each condition lasting 22s were mixed with 22s blocks of rest, during which no stimuli were presented and participants maintained fixation. Target and non-target words were distributed equally across conditions, and the target word in the HV condition was produced with equal probability by each of those four talkers.

3.1.4. fMRI data acquisition and analysis

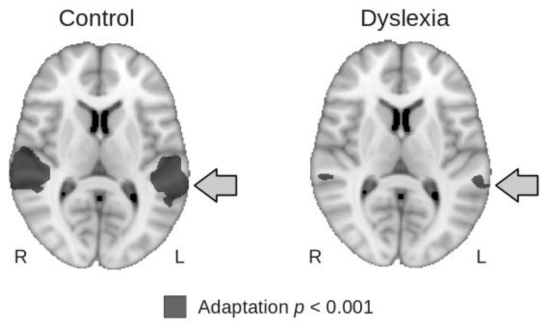
Data were acquired on a Siemens Trio 3T scanner with a 32-channel phased array head coil, including a high-resolution, T1-weighted multi-echo MPRAGE anatomical volume [acquisition parameters: TR=2350ms, TE=1.79ms, TI=1400ms, flip angle=7°, FOV=256×256, 176 slices, voxel resolution=1.0×1.0×1.0mm], and four functional runs containing 110 volumes each, collected using sparse-sampled T2*-weighted EPI scans [acquisition parameters: TR=5500ms, TA=2000ms, TE=30ms, flip angle=90°, voxel resolution= 3.125×3.125×4.0mm, FOV=64×64, 32 slices].

Cortical reconstruction of anatomical images was performed using Freesurfer v5.0.0. Functional data were analyzed in SPM8 using workflows in NiPyPE v0.3 (http://nipy.org/nipype), including rigid-body realignment for motion correction, volume smoothing ($6mm^3$), within-subject model design and estimation and contrast estimation. Functional data were coregistered to structural images using Freesurfer. Coregistered structural and functional images were aligned to a common space using nonlinear symmetric diffeomorphic mapping implemented in ANTS v1.5. Group contrasts were performed in SPM8.

3.2. Results & discussion

Brain regions engaged in talker normalization were those displaying neural adaptation, i.e., reduced activity in the single-talker condition compared to the multi-talker one. Control participants exhibited extensive adaptation to the repeated talker bilaterally in superior temporal gyrus, including primary auditory cortex and Wernicke's area. Participants with dyslexia exhibited only limited adaptation (Fig 2). Direct comparison between the two groups revealed significantly greater adaptation in the controls [t_{33}=2.45, p<0.01, cluster-corrected p<0.01], with the peak difference in left planum temporale – the auditory region central to phonological processing in speech perception [3].

Figure 2: Extensive neural adaptation revealed strong talker-normalization effects in controls, while such adaptation was extremely limited in dyslexic listeners.



The lack of robust neural adaptation to a repeated talker during a speech-perception task suggests dyslexic listeners may be intrinsically impaired in their ability to take advantage of talker normalization processes to facilitate speech perception. Instead, dyslexic listeners must rely on more physiologically expensive and inefficient exemplar-based strategies, even in the presence of predictable phonetic-phonemic correspondence for a single talker.

4. CONCLUSIONS

In two experiments that test talker-normalization processes, we find a pronounced, systematic deficit in dyslexic listeners' ability to take advantage of the increase in consistency and predictability of the phonetic-phonemic correspondence of a single talker that typically reduces the cognitive and physiological cost of speech perception for control listeners. These results explicate the link between aspects of speech perception (talker normalization processes and the abstract representations on which they rely) and the “phonological deficit” proposed in the clinical literature on reading impairment.

5. REFERENCES

[1] Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Rev.* 105, 251-279.

[2] Hazan, V., et al. 2009. Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *J. Speech Lang. Hear. Res.* 52, 1510-1529

[3] Hickok, G., Poeppel, D. 2007. The cortical organization of speech processing. *Nature Rev. Neurosci.* 8, 393-402.

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

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

[6] Ramus, F., Szenkovits, G. 2008. What phonological deficit? *Q. J. Exp. Psych.* 61, 129-141.