RAPID ADAPTATION TO TALKER-SPECIFIC PHONETIC DETAIL IS DISRUPTED BY NONINVASIVE BRAIN STIMULATION

Ja Young Choi[1,2] & Tyler K. Perrachione[2]

[1]Program in Speech and Hearing Bioscience and Technology, Harvard University, USA
[2]Department of Speech, Language, and Hearing Sciences, Boston University, USA
tkp@bu.edu

ABSTRACT

To resolve variation in acoustic-phonemic correspondences across talkers, listeners adapt to talkers' phonetic idiosyncrasies using preceding speech. Previous studies measured increased neural response in superior temporal lobe to talker variability, but it is unknown whether this region is causally involved in talker adaptation. We investigated how noninvasive brain stimulation affected talker adaptation during a speech processing task that factorially manipulated talker variability (single vs. mixed talkers) and speech context (isolated words vs. connected speech). In a between-subjects design, listeners received anodal, cathodal, or sham transcranial direct current stimulation of left superior temporal lobe while identifying target words. Connected speech reduced processing costs associated with mixed talkers; however, this effect was significantly attenuated under both anodal and cathodal stimulation compared to sham. Stimulation of left superior temporal lobe disrupts the brain's ability to use speech context to adapt to a talker, revealing this region's causal role in talker adaptation.

Keywords: speech perception; phonetic variability; talker adaptation; temporal lobe; tDCS

1. INTRODUCTION

In speech perception, listeners face the challenge of establishing correspondence between their abstract phonetic representations and the acoustic realizations of speech that are variable across talkers. When listeners encounter a new talker, they must ascertain a different acoustic-phonemic mapping from what they were using with the previous talker. This transition imposes additional processing costs, making listeners' speech perception slower or less accurate [15,17].

Natural speech tends to occur in a continuous stream rather than words or speech sounds in isolation, and listeners use information from the preceding speech context to accumulate talker-specific phonetic detail and rapidly adapt to a talker. As a result, preceding speech context not only biases the decision outcome of speech perception [9,14] but also reduces the processing costs associated with talker variability [2]. These results are consistent with models of speech perception that treat context as a frame of reference against which subsequent speech is compared [20] or a cue that narrows down the range of possible interpretations of an incoming signal [11].

The remarkably rapid time course of behavioral adaptation to talker-specific phonetic detail suggests that there must be processing mechanisms in the brain that can adapt on the order of seconds to new acoustic-phonetic mappings when the talker changes. Animal models of auditory plasticity show that response tuning in auditory neurons can adapt to new behaviourally-relevant sound statistics on this timescale [5,8], but how and where talker adaptation is achieved in the human brain remains unknown.

Neuroimaging studies have reported significantly more activation of superior temporal regions when listening to speech from mixed talkers than a single, consistent talker [1,21,23]. Bilateral superior temporal regions have also been implicated in processing the likelihood of phonetic category membership, with activity in these regions increasing as a function of the amount of phonetic variation [18]. Activation in this region may therefore reflect the processing cost associated with phonetic variability, but fMRI studies primarily reveal correlational, not causal, relationship between brain activity and behaviour [16].

We therefore aimed to investigate whether left superior temporal lobe causally underlies the brain's ability to use immediately preceding speech context to rapidly adapt to talkers on time scales on the order of one second. To determine the causal role of this region, we modulated cortical activity using high-definition transcranial direct current stimulation (HD-tDCS) – a safe, noninvasive technique that induces short-term changes in cortical excitability by employing weak electrical currents over the scalp [19]. The pattern of changes in listeners' behavior in response to modulated brain activity during speech processing tasks involving talker adaptation will provide causal evidence for whether left superior temporal lobe contributes to talker adaptation and on what timescale.

2. METHODS

2.1. Participants

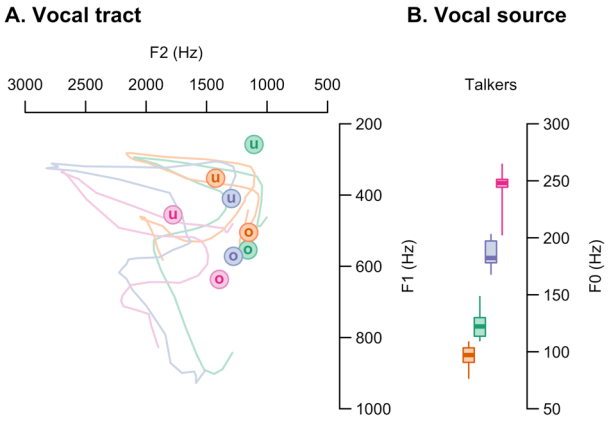
Native English-speaking adults (N = 60; 46 female,

14 male; age 18-31, M = 20.4 years) participated in this study. Participants had no metallic implants and no history of speech, language, hearing, or neurological disorder or significant head trauma. All participants were right-handed. Participants gave informed, written consent approved and overseen by the Institutional Review Board at Boston University.

2.2. Stimuli

Stimuli included two monophthongal target words "boot" and "boat", chosen because the acoustic-phonemic correspondence of the /u/-/o/ contrast is highly variable across talkers [7], imposing greater processing cost in a mixed-talker environment [3]. Target words were presented either in isolation or in connected speech, where they were preceded by the carrier phrase "I owe you a [boot/boat]." This carrier phrase was chosen because it provides an extensive sample of each talker's vowel space (Fig. 1A). Words and carrier phrases were recorded by two male and two female native American English speakers (Fig. 1B). The recordings were made in a sound-attenuated room sampling at 44.1kHz and 16bits. Connected speech sentences were synthesized by concatenating the carrier phrase to the target word, so that each talker's target word was identical in all conditions. Carrier phrases and target words were normalized to 65 dB SPL RMS amplitude in Praat.

Figure 1: Speech stimuli and phonetic variability. **(A)** Points labelled "u" and "o" indicate vowel formant frequencies in the target words; lines indicate the formant trace for each talker's carrier phrase ("I owe you a") in the connected speech condition. **(B)** Fundamental frequency of each talker's voice. Box plots show the distribution (median, interquartile range, extrema). Colors denote different talkers.

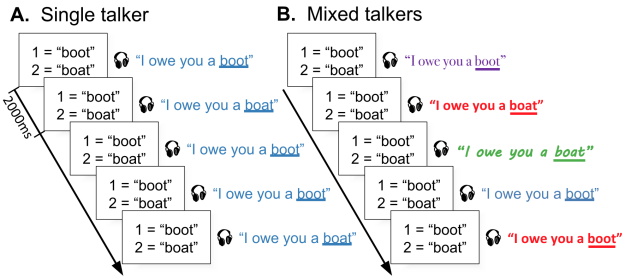


2.3. Behavioral task

Stimuli were presented in four blocks which factorially manipulated talker variability (*single-talker* vs. *mixed-talker*) and speech context (*isolated words* vs. *connected speech*). Each block consisted of 96 trials, each target word occurring in 48 trials per block. Stimulus presentation was pseudo-randomized such that the same word was not presented for more than three consecutive trials (Fig. 2). The order of conditions was counterbalanced across participants.

Participants were instructed to listen to the stimuli and identify the target word they heard as quickly and as accurately as possible by pressing the corresponding number on a keypad. Trials were presented at a rate of one per 2000ms. Stimulus delivery was controlled using PsychoPy v.1.8.1.

Figure 2: Behavioral task design. Participants identified words while listening to speech produced by either **(A)** a single talker or **(B)** mixed talkers. The *connected speech* condition is shown. Font/color combinations denote different talkers.



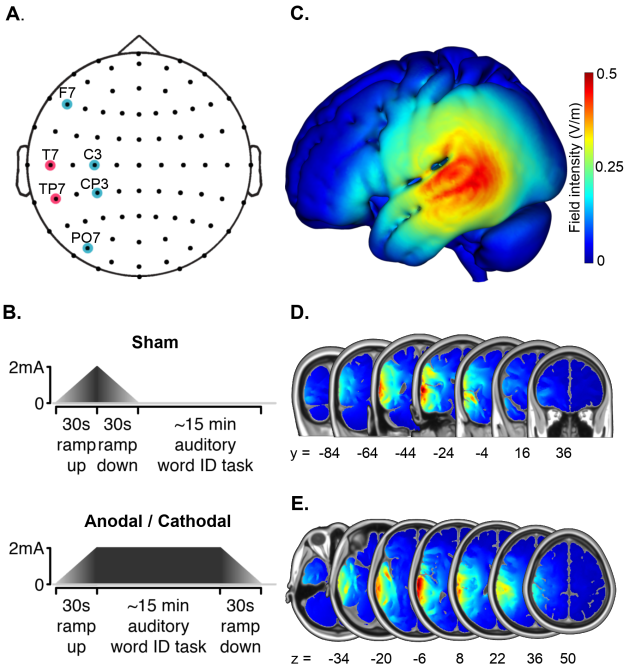
2.4. Transcranial direct current stimulation (tDCS)

In a between-subjects design, participants were randomly assigned to receive either sham (n = 20), anodal (n = 20) or cathodal (n = 20) tDCS during the task. Stimulation was applied using a Soterix M×N HD-tDCS system. Stimulating electrodes (cathodes for the cathodal condition, anodes for the anodal condition) were placed at electrode locations T7 and TP7 (in the 10-10 system [12]); return electrodes (anodes for the cathodal condition and cathodes for the anodal condition) were placed at C3, CP3, PO7 and F7 (Fig. 3A). This configuration, approximating the center-surround stimulation design optimal for achieving maximally focal stimulation intensity and current flow [4,13], was chosen to target left superior temporal cortex. Electrode locations were selected based on biophysical simulation of current flow in the human brain (Soterix HD-Explore, Soterix Medical, NY, USA). Peak estimated field intensity at the target location was 0.507 V/m (Fig. 3C,D,E).

For anodal and cathodal tDCS sessions, current was increased to the maximum stimulation intensity of 2 mA using a 30-s linear ramp after initiation (Fig. 3B). Stimulation magnitude remained at 2 mA for the entire duration of the task (~15 min), followed by a 30-s linear ramp-down at termination. For sham sessions, current was ramped up to 2 mA over 30 s and then immediately ramped down to 0 mA over 30 s,

where it remained for the entire duration of the task. Sham tDCS induces the initial mild dermal tingling sensation associated with tDCS without stimulating the brain areas during the task, keeping participants unaware of whether they were assigned to an active or sham stimulation. Electrode impedance was kept below 10 kΩ for all electrodes for all sessions.

Figure 3: tDCS paradigm. **(A)** Electrode configuration. Stimulating electrodes are shown in red; reference electrodes are shown in blue. **(B)** Schematic representation of current modulation during the experiment. Simulated current flow estimated by HD-Explore in **(C)** 3D view, **(D)** coronal view, and **(E)** axial view. The y- and z-coordinates refer to the slice location in MNI stereotaxic space.



2.5. Data analysis

Accuracy and response time data were analyzed for each participant in each condition. Accuracy was calculated as the proportion of trials in which the participant correctly identified the target words out of the total number of trials. Response times were log-transformed to approximate a normal distribution expected by the model. Only response times from correct trials were analyzed. Outlier trials deviating from the mean log response time in each condition by more than three standard deviations were excluded from analysis (< 1% of trials). Participants' response times were analyzed using linear mixed effects model with fixed factors including *speech context* (isolated vs. connected speech), *talker variability* (single- vs. mixed-talker), and *stimulation* (anodal vs. cathodal vs. sham), and with random effects including within-participants intercepts and slopes. Model fitting was computed based on maximum likelihood estimation using the packages *lme4* and *lmerTest* in R. Fixed factors' significance was determined by Type III analysis of variance (ANOVA). Significant effects from the ANOVA were followed by post-hoc pairwise contrasts on terms from the linear mixed effects model. Significance of main effects and interactions was determined by adopting significance criterion of $\alpha = 0.05$, with *p*-values based on the Sattertwaite approximation of the degrees of freedom.

3. RESULTS

Participants' word identification accuracy was at ceiling (98% ± 2%). Correspondingly, we were interested in speech processing *efficiency*, and the dependent measure for this study was response time (Table 1).

Table 1: Identification of target words in isolation. Mean ± s.d. response time (ms) in each stimulation and variability condition.

	Sham	Anodal	Cathodal
Single-Talker	745 ± 104	700 ± 76	717 ± 85
Mixed-Talker	836 ± 122	780 ± 87	805 ± 100
Difference	91 ± 66	79 ± 48	88 ± 82

Table 2: Identification of target words in connected speech. Mean ± s.d. response time (ms) in each stimulation and variability condition.

	Sham	Anodal	Cathodal
Single-Talker	679 ± 81	654 ± 75	645 ± 59
Mixed-Talker	708 ± 78	702 ± 79	697 ± 58
Difference	29 ± 49	48 ± 51	52 ± 49

3.1. Interference effects of talker variability

Response times in the single-talker conditions were significantly faster than the mixed-talker conditions under all three stimulation types (main effect of *talker variability*; $F(1,57) = 156.19$; $p \ll 0.001$). Response times in the connected-speech conditions were significantly faster than the isolated-word conditions (main effect of *speech context*; $F(1, 57) = 98.15$; $p \ll 0.001$).

We observed a significant *speech context* × *talker variability* interaction effect, such that listeners exhibited significantly more interference from talker variability when identifying words in isolation than in connected speech ($F(1, 22275) = 89.74$; $p \ll 0.001$).

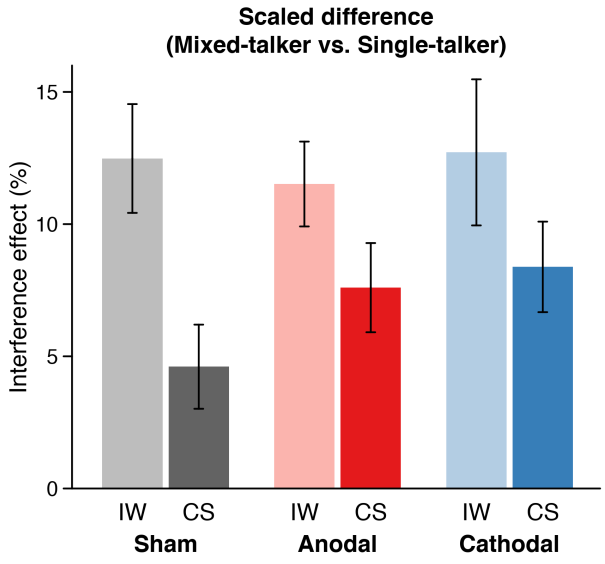
3.2. Effects of neurostimulation on talker adaptation

Stimluation had no significant effect on overall reaction times ($F(2, 57) = 1.03$; $p = 0.36$). However, there

was a significant *stimulation* × *speech context* × *talker variability* interaction ($F(2, 22275) = 10.66$; $p < 0.01$), indicating that the amount of interference imposed by processing mixed talkers (vs. a single talker) during isolated words (vs. connected speech) differed across the three stimulation conditions. Pairwise contrasts on the three-way interaction terms of the linear mixed-effects model revealed that the talker-variability-by-speech-context interaction was significantly greater under sham than either anodal stimulation ($\beta = 0.0038$, $SE = 0.0013$, $t = 2.97$, $p < 0.01$) or cathodal stimulation ($\beta = 0.0034$, $SE = 0.0013$, $t = 2.65$, $p < 0.01$). These three-way interactions reveal that the facilitatory effect that connected speech usually has on reducing the interference effect of processing mixed-talker speech was diminished under active stimulation compared to sham (Fig. 4).

There was no significant effect of stimulation in the isolated words conditions alone (*sham* vs. *anodal* $\beta = 0.014$, $SE = 0.018$, $t = 0.76$, $p = 0.45$; *sham* vs. *cathodal* $\beta = 0.0023$, $SE = 0.018$, $t = 0.13$, $p = 0.90$).

Figure 4: Mean interference effects of talker variability for isolated words (IW) and connected speech (CS) in each stimulation condition. Error bars indicate standard error of mean across participants. The interference effect is calculated as the scaled difference between the average response time in mixed-talker condition and the single-talker condition: 100 × [(mixed-talker) – (single-talker)] / (single-talker).



4. DISCUSSION

We investigated how noninvasive stimulation of left superior temporal lobe influences the brain's ability to use preceding speech context to rapidly adapt to talkers. Compared to sham, both anodal and cathodal stimulation disrupted rapid talker adaptation in connected speech. However, there was no difference in the interference effect among the three different stimulation types when processing isolated words, suggesting that disruption of neurocomputational processes in left superior temporal lobe impairs the brain's ability to rapidly adapt to a talker, but not its ability to adapt over longer timescales.

Previous fMRI studies have reported neural adaptation effects under tasks similar to our isolated-word condition: reduced activation of superior temporal areas is found in single-talker blocks relative to mixed-talker blocks [21,23]. Extending upon these findings, we observed that the reduction of the interference effect in connected speech but not isolated words was attenuated by stimulation of left superior temporal lobe. This result suggests that left superior temporal lobe is causally involved in rapid integration of context information into the perceptual system during connected speech. Thus, early integration of talker and speech information [10] likely occurs in this region, where neural response differences between single- and mixed-talker speech likely reflects the additional neurocomputational demands deployed to process talker variability.

Left hemisphere tDCS only affected rapid talker adaptation in connected speech, perhaps due to hemispheric differences in temporal integration of speech information [24]. Future work will need to explore whether and how tDCS of right superior temporal lobe disrupts adaptation and on what timescale, and what consequence, if any, stimulation of non-auditory areas (e.g., prefrontal cortex) has on talker adaptation.

We also found no difference in the behavioral effect of stimulation between anodal and cathodal polarities, which are thought to increase and decrease cortical excitability, respectively [19]. Although their behavioral effects were similar, the mechanism by which anodal and cathodal tDCS disrupt talker adaptation may differ: anodal stimulation may reduce balanced precision between excitation and inhibition underlying neocortical adaptation [22] resulting in less precise re-tuning, while cathodal stimulation may reduce the magnitude of short-term plastic changes, making them less specific [6].

Our findings demonstrate the potential for tDCS as a research tool for exploring the functional neuroanatomy of speech perception, expanding its current usage in research on higher-level language processing. While other neurostimulation technologies, including transcranial magnetic stimulation (TMS) have been used in speech research, tDCS offers an advantage as it does not produce the loud acoustic noise resulting from the magnetic coils in TMS. tDCS can therefore further be used to better understand the functional neuroanatomical bases of communication disorders such as dyslexia, where neural dysfunction in rapid speech adaptation has been reported [21].

5. REFERENCES

[1] Belin, P., Zatorre, R.J. 2003. Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, 14, 2105–2109.

[2] Choi, J.Y., Perrachione, T.K. Time and information in perceptual adaptation to speech. Submitted.

[3] Choi, J.Y., Hu, E.R., Perrachione, T.K. 2018. Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attn. Percept. Psychophys.* 80, 784–797.

[4] Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., Bikson, M. 2009. Gyri-precise head model of transcranial direct current stimulation: improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimul.*, 2, 201–207.

[5] Fritz, J., Shamma, S., Elhilali, M., Klein, D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.*, 6, 1216–1223.

[6] Froemke, R.C., Merzenich, M.M., Schreiner, C.E. 2007. A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450, 425–429.

[7] Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099–3111.

[8] Jääskeläinen, I.P., Ahveninen, J., Belliveau, J.W., Raij, T., Sams, M. 2007. Short-term plasticity in auditory cognition. *Trends Neurosci.*, 30, 653–661.

[9] Johnson, K. 1990. The role of perceived speaker identity in F0 normalization of vowels. *J. Acoust. Soc. Am.* 88, 642–654.

[10] Kaganovich, N., Francis, A.L., Melara, R.D. 2006. Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Res.* 1114, 161–172.

[11] Kleinschmidt, D.F., Jaeger, T.F. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* 122, 148–203.

[12] Klem, G.H., Lüders, H.O., Jasper, H.H., Elger, C., 1999. The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol*. 52, 3–6.

[13] Kuo, H.I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M.F., Nitsche, M.A. 2013. Comparing cortical plasticity induced by conventional and high-definition 4 × 1 ring tDCS: A neurophysiological study. *Brain Stimul.*, 6, 644–648.

[14] Ladefoged, P., Broadbent, D.E. 1957. Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98–104.

[15] Magnuson, J.S., Nusbaum, H.C. 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Human*, 33, 391–409.

[16] Mather, M., Cacioppo, J.T., Kanwisher, N. 2013. How fMRI can inform cognitive theories. *Perspect. Psychol. Sci.*, 8, 108–113.

[17] Mullennix, J.W., Pisoni, D.B. 1990. Stimulus variability and processing dependencies in speech perception. *Percept. Psychophys.* 47, 379–390.

[18] Myers, E.B. 2007. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, 45, 1463–1473.

[19] Nitsche, M.A., Paulus, W. 2000. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639.

[20] Nusbaum, H.C., Morin, T.M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech perception, production, and linguistic structure* . Tokyo: Ohmasha Publishing, 113–134.

[21] Perrachione, T.K., Del Tufo, S.N., Winter, R., Murtagh, J., Cyr, A., Chang, P., Halverson, K., Ghosh, S.S., Christodoulou, J.A., Gabrieli, J.D.E. 2016. Dysfunction of rapid neural adaptation in dyslexia. *Neuron*, 92, 1383–1397.

[22] Wehr, M., Zador, A.M., 2003. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426, 442–446.

[23] Wong, P.C.M., Nusbaum, H.C., Small, S.L. 2004. Neural bases of talker normalization. *J. Cog. Neuro*. 16, 1173–1184.

[24] Zatorre R.J., Belin, P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb. Cort.* 11, 946–953.