

Noninvasive neurostimulation of left temporal lobe disrupts rapid talker adaptation in speech processing

Ja Young Choi^{1,2} & Tyler K. Perrachione^{1*}

¹ Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts

² Program in Speech and Hearing Bioscience and Technology, Harvard University, Cambridge Massachusetts

***Correspondence:**

Tyler K. Perrachione, Ph.D.
Department of Speech, Language, and Hearing Sciences
Boston University
635 Commonwealth Ave.
Boston, MA 02215
Phone: +1.617.358.7410
Email: tkp@bu.edu

Abstract

Talker adaptation improves speech processing efficiency by reducing possible mappings between talkers’ speech acoustics and listeners’ phonemic representations. We investigated the functional neuroanatomy of talker adaptation by applying noninvasive neurostimulation (high-definition transcranial direct current stimulation; HD-tDCS) to left superior temporal lobe while participants performed an auditory word identification task. We factorially manipulated talker variability (single vs. mixed talkers) and speech context (isolated words vs. connected speech), measuring listeners’ speech processing efficiency under anodal, cathodal, or sham stimulation. Speech processing was faster for single talkers than mixed talkers, and connected speech reduced the additional processing costs associated with mixed-talker speech. However, the beneficial effect of connected speech in the mixed-talker condition was significantly attenuated under both anodal and cathodal stimulation versus sham. Stimulation of left superior temporal lobe disrupts the brain’s ability to use local phonetic context to rapidly adapt to a talker, revealing this region’s causal role in talker adaptation.

Keywords: speech perception; tDCS; adaptation; phonetic variability; superior temporal cortex

1. Introduction

Mapping acoustic speech signals onto abstract phonemic representations is a key challenge in speech perception, as the acoustic realization of speech varies substantially across talkers. Thus, when listeners encounter a new talker, they need to quickly ascertain the acoustic-phonemic mappings that correspond to that talker, resulting in an additional processing cost relative to when the talker does not change (Johnson, 2005). The additional processing costs incurred by talker variability have been extensively shown in previous behavioral studies, in which listeners' performance in speech perception tasks gets slower or less accurate when they listen to mixed talkers rather than a single talker (Choi, Hu, & Perrachione, 2018; Strange et al., 1976; Assman, Nearey, & Hogan, 1982; Green, Tomiak, & Kuhl, 1997; Mullennix & Pisoni, 1990; Magnuson & Nusbaum, 2007). Correspondingly, neuroimaging studies have routinely shown that listening to speech from mixed talkers leads to greater activation of superior temporal cortices compared to listening to speech from a single talker (Belin & Zatorre, 2003; Wong, Nusbaum & Small, 2004; Chandrasekaran et al., 2011; Perrachione et al., 2016; Zhang et al., 2016).

One mechanism by which listeners adapt to a talker is by using the immediately preceding speech context (Johnson 1990; Nearey, 1989). Speech in real life almost always occurs in a continuous stream, rather than a word or a speech sound in isolation, and previous speech sounds produced by a talker provide listeners with contextual information about the phonetic space of that talker. Previous studies have shown that preceding speech context biases the decision outcome of speech perception (Ladefoged & Broadbent, 1957; Johnson, 1990) and reduces the processing costs associated with talker variability (Choi & Perrachione, 2019).

These empirical results lend support to several related models of speech processing that account for how contextual information is integrated by the perceptual system. *Contextual tuning theory* treats preceding context as a frame of reference against which the following speech is compared (Nusbaum & Morin, 1992). Under this model, listeners use information embedded in the first speech sounds produced by a new voice to build an internal representation of the vocal tract (i.e., formant space) specific to the talker, which is then used to interpret following speech sounds produced by the same voice. Building upon this theory, Magnuson & Nusbaum (2007) proposed that speech perception is an *active control process*, in which listeners build hypotheses regarding the interpretation of incoming signals and check them against the speech sounds that they encounter. This process is proposed to be triggered when listeners detect a change of talker and to operate until a stable mapping between the speech sounds produced by the new talker and the listeners' internal phonetic categories is established. In an alternative framework, episodic models of speech perception (e.g., Goldinger, 1998) also highlight the role of previously encountered speech in processing subsequent speech signals. Recently formalized as the *ideal adapter framework*, this model posits that listeners use cues prior to a speech target to narrow down the range of possible interpretations of incoming speech based on prior experiences with an individual or class of speakers (Kleinschmidt & Jaeger, 2015).

Despite the theoretical and empirical work on rapid talker adaptation using context, the neural mechanisms of talker adaptation still remain elusive. Talker variability is consistently found to increase neural activation in superior temporal lobe (Wong et al., 2004; Belin & Zatorre, 2003; Chandrasekaran et al., 2011; Perrachione et al., 2016), but the causal contribution of this region to processing talker variability is still unknown. Animal models of auditory cortical dynamics and plasticity have elaborated the processes by which neural representations of behaviorally-relevant sounds can be tuned by context over short time-scales on the order of seconds (Fritz et al., 2003; Froemke et al., 2007; Herrmann et al., 2015). Similar

mechanisms may constitute the neurobiological basis for talker adaptation during speech perception by human listeners, but a synapse- or circuit-level understanding of adaptation in speech processing remains beyond the abilities of current human systems neuroscience research. However, a means for studying the causal contribution of larger brain structures in processing talker variability is possible through noninvasive brain stimulation. Transcranial direct current stimulation (tDCS) is a safe, noninvasive technique that modulates cortical excitability and plasticity by employing weak electric currents over the scalp, with anodal stimulation increasing cortical excitability and cathodal stimulation decreasing it (Nitsche & Paulus, 2000). Thus, causal evidence for the involvement of a particular brain area in processing talker variability can be inferred if targeted stimulation of that region results in behavioral changes in speech processing, and the direction and degree of change associated with each polarity of stimulation can better inform us of circuit-level understanding of talker adaptation.

In this study, we aimed to investigate whether the left superior temporal lobe causally underlies the brain's ability to adapt to talkers and, if so, the timescale of its involvement in talker adaptation. While previous neuroimaging studies have shown that processing speech from multiple talkers vs. a single talker elicits greater response in bilateral superior temporal regions, the source of this increased activation may differ between the two hemispheres: Compared to a single-talker condition, a mixed-talker condition increases not only phonetic variability but also variability in the source of speech (i.e., talker identity). Several studies have specifically contrasted processing talker identity vs. speech content, and have consistently found left-lateralized processing of the verbal content in speech and right-lateralized processing of voice content (e.g., Stevens, 2004; von Kriegstein et al., 2003). These results are consistent with the classic finding that phonological processing of speech is mediated by the left hemisphere (Wernicke, 1874; Scott et al., 2000; Liebenthal et al., 2005; Obleser et al., 2007).

In a mixed between/within-subjects design, participants were assigned to groups receiving either anodal, cathodal, or sham high-definition (HD) tDCS to left superior temporal lobe while performing a word identification task. All listeners identified which of two phonetically-confusable target words they heard ("boot" or "boat") while we factorially varied talker variability (single vs. mixed talkers) and speech context (isolated words vs. connected speech). Using participants' response time to the target word as our dependent variable, we focused on how the speed of word identification changes as a consequence of listening to mixed talkers as opposed to a single talker, and how that difference varies as a function of speech context. This allowed us to explore talker adaptation at two different timescales – within each block (on the order of seconds) and within each trial (on the order of hundreds of milliseconds). Comparing the response time differences between different stimulation groups, we investigated how noninvasive stimulation of left superior temporal lobe influenced talker adaptation.

We expected to replicate the interference effect of talker variability, that response times are slower for mixed- vs. single-talker speech (Mullennix & Pisoni, 1990; Choi, Hu, & Perrachione, 2018), and to replicate the finding that extrinsic talker adaptation leads to a smaller interference effect in connected speech vs. isolated words (Choi & Perrachione, 2019). We expected that anodal stimulation would facilitate talker adaptation, whereas cathodal stimulation would interfere with the process, as anodal stimulation of the left temporal region in healthy individuals has often been shown to improve performance in speech and language domain (reviewed in Zoefel & Davis, 2017). However, it is important to note that the heuristic hypothesis that anodal stimulation enhances, while cathodal stimulation impairs, a target behavior does not necessarily reflect the complex neurobiological mechanisms that electrical stimulation of the cortex affects (Dayan et al., 2013; Bestmann et al., 2015). Finally, we hypothesized that stimulation would

affect talker adaptation for connected speech vs. isolated words differently, given the unique role of the left hemisphere in processing connected speech (Peelle, 2012).

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 as indicated by the Edinburgh Handedness Inventory (Oldfield, 1971). Participants gave informed, written consent approved and overseen by the Institutional Review Board at Boston University.

2.2. Stimuli

Stimuli included two target words “boot” and “boat.” We chose these words because the acoustic-phonemic correspondence of the /u/-/o/ contrast is highly talker-dependent; the acoustic realization of the vowels /u/ and /o/ exhibits extensive overlap across talkers that listeners must resolve on a talker-specific basis to correctly identify the target phoneme (Hillenbrand et al., 1995) and therefore imposes greater processing interference in a mixed-talker environment (Choi et al., 2018). 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**), offering listeners talker-specific phonetic details that they can use to calibrate their perception of the following vowel in the following target word (Joos, 1948; Nusbaum & Morin, 1992; Johnson, 1990). Words and carrier phrases were recorded by two male and two female native speakers of American English (**Fig. 1A**). The recordings were made in a sound-attenuated room with a Shure MX153 earset microphone and Roland Quad Capture sound card sampling at 44.1kHz and 16bits. Among numerous tokens from these speakers, the recordings in which the boot / boat distinction was most evident based on their formant frequencies – and which were least dissimilar in noncontrastive features such as voice pitch, amplitude envelope, and duration – were chosen as the final stimulus set. The mean duration of the target words was 228 ms (range: 203-256 ms), and the mean duration of the prepended carrier phrases was 609 ms (range: 543-656 ms). Connected speech sentences were synthesized by concatenating the naturally-recorded carrier phrase to the target word, so that the same target word stimuli from each talker were used in all conditions. Carrier phrases and target words were normalized to 65 dB SPL RMS amplitude in Praat (Boersma, 2001). For stimuli and stimulus-delivery scripts, see **Open-Source Dataset** section.

2.3. Behavioral task

Participants’ task on each trial was to listen to the stimulus and indicate whether they heard “boot” or “boat” as quickly and accurately as possible by pressing the corresponding number on the keypad. Trials were organized into four blocks that factorially manipulated *talker variability* (single-talker vs. mixed-talker) and *speech context* (isolated words vs. connected speech), with each block corresponding to one of the four conditions. Each block consisted of 96 trials, with each target word occurring in 48 trials per block. Stimulus presentation was pseudo-randomized such that the same target word was not presented for more than three consecutive trials (**Fig. 1B**). The order of conditions was counterbalanced across participants using Latin-square permutations. For each participant, the same talker served as the single talker

in both single-talker blocks, and which of the four talkers was used in the single-talker conditions was counterbalanced across participants.

The duration of each trial, including the duration of the stimulus and the time for participants to respond with the keypad, was kept at 2000ms across all conditions. Stimulus delivery was controlled using PsychoPy v.1.8.1 (Peirce, 2007). The total experiment duration was approximately 13 minutes.

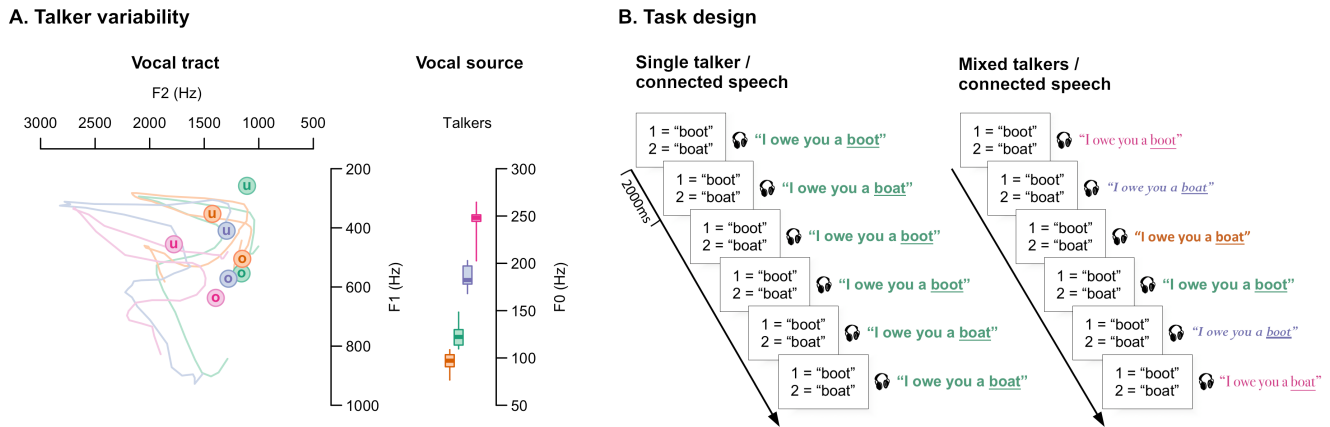


Figure 1: Stimulus variability across talkers and task design. (A) Phonetic variability of stimuli across talkers. Left: F1 and F2 of the target words (circles; /u/ boot, /o/ boat) and the F1-F2 trajectory of the carrier phrase (lines; "I owe you a"). Right: f_0 (vocal pitch) distribution for all talkers' recordings. Colors denote different talkers. (B) Behavioral task design. Participants identified words while listening to speech produced by either a single talker (left) or mixed talkers (right). The connected speech conditions are shown. Font/color combinations denote different talkers.

2.4. High-definition transcranial direct current stimulation (HD-tDCS)

In a between-subjects design, participants were randomly assigned to receive either sham ($n = 20$), anodal ($n = 20$) or cathodal ($n = 20$) HD-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 (Klem et al., 1999); return electrodes (anodes for the cathodal condition and cathodes for the anodal condition) were placed at C3, CP3, PO7 and F7 (Fig. 2A). This configuration, which approximates the center-surround stimulation design that has been shown to be optimal for achieving maximally focal stimulation intensity and current flow (Datta et al., 2009; Kuo et al., 2013), was chosen to focally 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. 2B-D).

For anodal and cathodal HD-tDCS sessions, current was increased to the maximum stimulation intensity of 2 mA using a 30-s linear ramp after initiation. Stimulation magnitude remained at 2 mA for the entire duration of the task (~13 min), followed by a 30-s linear ramp-down at termination. For sham HD-tDCS sessions, current was linearly ramped up to 2 mA over 30 s and then immediately ramped back down to 0 mA over 30 s, where it remained for the entire duration of the task. Sham HD-tDCS induces the initial mild dermal tingling sensation associated with HD-tDCS without stimulating the brain areas below the electrodes during the task, thus keeping participants unaware as to whether they were assigned to an active stimulation or sham control condition. Participants were asked to fill out a questionnaire after

completing the experiment to ensure that HD-tDCS did not cause excessive discomfort. Electrode resistance was kept below 10 k Ω for all electrodes for all sessions.

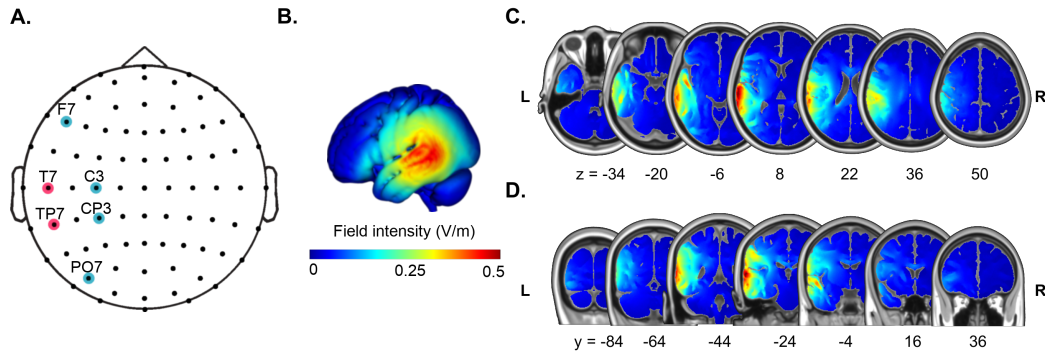


Figure 2: tDCS paradigm. (A) Electrode configuration. Stimulating electrodes are shown in red; reference electrodes are shown in blue. Simulated current flow estimated by HD-Explore in (B) 3D view, (C) axial view, and (D) coronal view. The y - and z -coordinates refer to the slice location in MNI stereotaxic space. Slices are shown in neurological convention.

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 more closely 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 a linear mixed effects model with fixed factors including *speech context* (isolated words 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 within-participant slopes for the effects of context and variability. Significance of factors was determined in a Type III analysis of variance (ANOVA). Significant effects from the ANOVA were followed by post-hoc pairwise analyses by testing contrasts on the terms in the linear mixed effects model using the package *lmerTest* in R. Contrasts were treatment-coded, with baseline levels of isolated words (*speech context*), single-talker (*talker variability*), and sham (*stimulation*). Significance of main effects and interactions was determined by adopting the significance criterion of $\alpha = 0.05$, with p -values based on the Satterthwaite approximation of the degrees of freedom.

3. Results

Participants' word identification accuracy was at ceiling ($98\% \pm 2\%$), with no effect of stimulation condition on participants' accuracy. As this study was primarily designed to investigate speech processing efficiency, the principal dependent measure was response time (**Table 1**).

In the post-experiment questionnaire, the number of participants who reported scalp sensations related to HD-tDCS did not differ between sham and active (combined anodal and cathodal) stimulation groups ($\chi^2(1) = 1.68, p = 0.19$). Participants reported mild to moderate tingling (84% of all participants); mild pain (36%), and mild burning sensations (29%). The number of participants reporting each type of sensation did not differ between sham and active stimulation groups (tingling $\chi^2(1) = 0.19, p = 0.67$; pain

$\chi^2(1) = 1.49, p = 0.22$; burning $\chi^2(1) = 0.22, p = 0.64$). The lack of group difference in these responses suggests that participants were effectively blinded as to whether they received active or sham stimulation.

Table 1. Mean \pm s.d. response time (ms) in each condition

	Sham		Anodal		Cathodal	
	Isolated Words	Connected Speech	Isolated Words	Connected Speech	Isolated Words	Connected Speech
Single-Talker	745 \pm 104	679 \pm 81	700 \pm 76	654 \pm 75	717 \pm 85	645 \pm 59
Mixed-Talker	836 \pm 122	708 \pm 78	780 \pm 87	702 \pm 79	805 \pm 100	697 \pm 58
Difference	91 \pm 66	29 \pm 49	79 \pm 48	48 \pm 51	88 \pm 82	52 \pm 49

3.1. Interference effects of talker variability

The ANOVA of the linear mixed-effects model revealed a robust main effect of *talker variability* ($F(1, 57) = 156.19; p \ll 0.0001$), showing that response times in the mixed-talker conditions were significantly slower than the single-talker conditions overall. Response times in the connected-speech conditions were also significantly faster overall compared to the isolated-word conditions (main effect of *speech context*; $F(1, 57) = 98.15; p \ll 0.0001$).

We observed a significant *speech context* \times *talker variability* interaction effect ($F(1, 22275) = 89.74; p \ll 0.0001$), indicating that the magnitude of processing interference from the mixed-talker condition differed depending on whether the target words were embedded in continuous speech or presented in isolation. Listeners exhibited significantly more interference from talker variability when recognizing words in isolation than in connected speech.

3.2. Effects of neurostimulation on talker adaptation

The HD-tDCS manipulation did not have a significant effect on overall response time (no main effect of *stimulation*; $F(2, 57) = 1.03; p = 0.36$). There was also no significant *stimulation* \times *talker variability* interaction ($F(2, 57) = 0.40; p = 0.67$), nor *stimulation* \times *speech context* interaction ($F(2, 57) = 1.14; p = 0.33$).

Critically, there was a significant *stimulation* \times *speech context* \times *talker variability* interaction ($F(2, 22275) = 5.33; p < 0.01$), indicating that the amount of benefit obtained from connected speech under talker variability differed among the three stimulation conditions (**Fig. 3**). To understand the three-way interaction across three levels of the *stimulation* factor, we turned to the pairwise contrasts on the three-way interaction terms of the linear model: The *talker variability* \times *speech context* \times *stimulation* interaction was significant for anodal vs. sham ($\beta = 0.0038, SE = 0.0013, t = 2.97, p < 0.01$) and cathodal vs. sham ($\beta = 0.0034, SE = 0.0013, t = 2.65, p < 0.01$) stimulation. This indicates that the effect of connected speech on mitigating the interference effect of mixed talkers was smaller under anodal and cathodal stimulation conditions than under sham stimulation. Furthermore, in models on subsets of the data examining only the single- and mixed-talker conditions separately, the *stimulation* \times *speech context* interaction effect for mixed talkers was nearly three times larger than the respective effect for a single talker (Anodal: $\beta_{\text{interact.}} = 0.06$ (50 ms) vs. 0.02 (20 ms); Cathodal: $\beta_{\text{interact.}} = 0.03$ (20 ms) vs. 0.01 (-6 ms)). That is, compared to sham, HD-tDCS disrupts the brain's ability use the immediately preceding speech context to rapidly adapt to each new talker in a mixed-talker context.

In the isolated words condition alone, however, the magnitude of the talker variability effect (mixed vs. single talkers) was not affected by either of the active stimulation conditions compared to sham (the contrast on the *stimulation* \times *talker variability* interaction term for isolated words only; sham vs. anodal $\beta = 0.014$, $SE = 0.018$, $t = 0.76$, $p = 0.45$; sham vs. cathodal $\beta = 0.0023$, $SE = 0.018$, $t = 0.13$, $p = 0.90$).

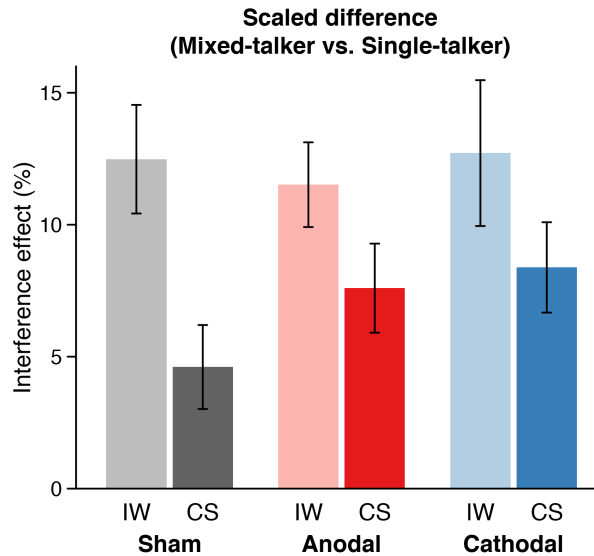


Figure 3: Processing cost of talker variability by speech context and stimulation condition. Mean interference effects of talker variability for isolated words (IW) and connected speech (CS) in each stimulation condition. Taller bars reflect greater differences in word identification response time between the mixed- vs. single-talker conditions. The *interference effect of talker variability* is calculated as the scaled difference between the average response time (RT) in mixed-talker condition and the single-talker condition: $100 \times [(RT_{\text{mixed}}) - (RT_{\text{single}})] / (RT_{\text{single}})$. Error bars indicate standard error of mean across participants.

4. Discussion

In this study, we used noninvasive neurostimulation to investigate the causal role of left superior temporal lobe in talker adaptation. We observed a significant interaction with stimulation such that, compared to sham, both anodal and cathodal stimulation disrupted rapid talker adaptation in connected speech. When processing isolated words, however, the three different types of stimulation did not differentially affect processing efficiency between single- and mixed-talker speech. These results raise the possibility that there is a dissociation between two timescales of—or mechanisms for—adaptation to a talker using preceding speech context, in which disruption of neurocomputational processes in left superior temporal lobe impairs the brain’s ability to rapidly adapt to a talker on a timescale as short as within a sentence (< 1 s), but not its ability to adapt over longer timescales.

4.1. Causal involvement of left superior temporal region in rapid talker adaptation

Our observations extend previous fMRI studies that have reported reduced activation in superior temporal areas in single-talker blocks relative to mixed-talker blocks (i.e., neural adaptation effects) when subjects performed tasks similar to our isolated-word condition (Wong et al., 2004; Belin & Zatorre, 2003; Chandrasekaran et al., 2011; Perrachione et al., 2016; Zhang et al., 2016). In addition to the correlation

between speech processing behavior and neural activity established by those previous neuroimaging studies, we found that the extent to which connected speech can offset the interference effect of mixed talkers was disrupted by electrical stimulation of left superior temporal lobe. This result appears to be specific to rapid integration of context information during talker adaptation from connected speech rather than a more general effect on speech processing efficiency. Increasing (or decreasing) cortical excitability of left superior temporal lobe via noninvasive neurostimulation did not generally speed up (or slow down) speech processing, neither overall nor in either speech context separately. This pattern of results suggests that left superior temporal lobe is causally involved in rapid integration of context information during connected speech. Thus, the early integration of talker and speech information likely occurs in this structure, where neural response differences between single- and mixed-talker speech likely reflect the additional computational demands in processing talker variability (Kaganovich et al., 2006).

In two conditions of this study, we preceded the target words with a carrier phrase to provide listeners with talker-specific vocal and phonetic details, giving them an extrinsic context from which they could develop expectations about the correspondence between speech acoustics and phonemic categories (Johnson, 1990; Nusbaum & Morin, 1992; Magnuson & Nusbaum, 2007). Auditory expectations sharpen neural responses to relevant stimulus features (Fritz et al., 2003; Todorovic et al., 2011), which may underlie our behavioral outcomes showing overall faster response times when the target words were preceded by an adapting carrier phrase.

Although we specifically operationalized the resolution of acoustic-phonemic ambiguity through differences in speech phonetics across talkers, it is possible that the computations carried out by superior temporal lobe may contribute to resolving phonetic ambiguity more generally. For instance, in a phonetic category judgment task, recruitment of superior temporal lobe bilaterally is greater when listeners are less certain of phonetic category membership (Myers, 2007), suggesting that this region may be the locus of resolving variable acoustic-phonemic mappings even when the source of variability is not related to differences across talkers. However, neuroimaging studies of processing variability in speech perception have also almost exclusively operationalized speech variability as phonetic variability between talkers, and future work must ascertain whether analogous normalizing processes also underlie within-talker variation arising from, for example, speech rate or coarticulation.

These results also broach the question of whether talker adaptation comprises neurocomputational processes specific to speech processing or reflects a more domain-general phenomenon underlying auditory adaptation. Even non-speech extrinsic contexts have been shown to affect speech processing in a manner similar to talker adaptation (Sjerps, Mitterer, & McQueen, 2011; Laing et al., 2012), demonstrating that auditory perceptual adaptation to speech during talker adaptation may actually be occurring via more fundamental auditory processes underlying stimulus adaptation (Herrmann et al., 2015). Correspondingly, as we discuss below, stimulation of left superior temporal lobe may ultimately be affecting feedforward adaptation of auditory circuits, rather than computations specific to speech processing.

4.2. Effects independent of stimulation polarity

Behaviorally, there was no difference in the effect of stimulation between anodal and cathodal polarities, which are thought to increase and decrease cortical excitability, respectively (Nitsche & Paulus, 2000). This may be due to the fact that rapid re-tuning of auditory perception relies on the precise (re-)balancing between excitatory and inhibitory activity, rather than a unidirectional process. Although the behavioral effects of the two polarities were similar, the mechanism by which HD-tDCS disrupts talker

adaptation may nonetheless differ: anodal stimulation may reduce the balanced precision between excitation and inhibition that underlies neocortical adaptation (Wehr & Zador, 2003), resulting in less precise re-tuning and thereby reducing perceptual efficiency. Cathodal stimulation, meanwhile, may reduce the magnitude of short-term changes to synaptic weights (Froemke et al., 2007), making them less specific. Application of a small electric current over the scalp, anodal or cathodal, may have interrupted this balance between excitation and inhibition in different ways, thus degrading the facilitatory effect of feedforward stimulus continuity on perception. Moreover, as the effect of HD-tDCS varies depending on various factors such as simultaneity between stimulation and the task, stimulation magnitude and duration, electrode placements, and cognitive load (Ohn et al., 2008; Roe et al., 2016; Thair et al., 2017), HD-tDCS polarity effects in cognitive domains cannot simply be reduced to an “anodal-excitation and cathodal-inhibition” effect heuristic (Jacobson et al., 2012). Indeed, both anodal and cathodal stimulation of auditory cortex have been shown to increase the magnitude of various auditory evoked potentials (Zaehle et al., 2011).

4.3. No effect of stimulation on talker adaptation to isolated words

In single-talker blocks, listeners can benefit from using the same talker-specific acoustic-phonemic mappings on every trial, even when they are listening to isolated words. When there is context that immediately precedes the target words, the processing costs associated with mixed-talker speech are reduced, because listeners can rapidly ascertain some talker-specific cues from the local context, even when the talker differs from the previous trial (Choi & Perrachione, 2019). By using both the isolated-word and connected speech conditions in this experiment, we were able to investigate how the left superior temporal region is involved in talker adaptation on varying timescales.

Our study showed that anodal and cathodal stimulation of left superior temporal lobe reduced the benefit of adaptation on short timescales (i.e., for connected speech) but did not reduce the adaptation effect on longer timescales (i.e., for isolated words). Since neurostimulation revealed no causal role of left superior temporal region in talker adaptation in the scale of seconds, such adaptation may be mediated by other brain regions. In addition to the superior temporal lobe, Wong and colleagues (2004) found talker adaptation-related activation in superior parietal lobe. They suggested activation in this region may reflect the additional cognitive effort demanded by constant attentional reorientation to new talkers in mixed-talker blocks. Future work will need to assess whether applying noninvasive neurostimulation to superior parietal lobe will affect talker adaptation to isolated words, as predicted by the attentional-reorientation hypothesis.

That left hemisphere stimulation did not affect talker adaptation from isolated words may also be due to hemispheric differences in temporal integration of connected vs. unconnected speech information. For instance, Peelle (2012) advances the idea that differences in left-lateralized vs. bilateral responses to speech depend primarily on whether speech is encountered in a connected (i.e., phrasal or sentential) vs. unconnected (i.e., individual words or syllables) context. Such a framework is consistent with our results, where left hemisphere HD-tDCS disrupted talker adaptation in a connected speech context, but not in an isolated word context, where the right hemisphere’s putative role in processing unconnected speech was undisrupted. This pattern of results is also consistent with a longstanding supposition that the two cerebral hemispheres may be involved in integrating auditory information on different timescales (e.g., Zatorre & Belin, 2001; Boemio et al., 2004; Abrams et al., 2008), notwithstanding what those particular timescales may be. Future work is thus clearly needed to explore how HD-tDCS of right superior temporal lobe also affects talker adaptation, and whether it does so for connected vs. unconnected speech contexts.

4.4. Limitations and future directions

In our application of HD-tDCS to left superior temporal cortex, we observed a three-way interaction between stimulation, talker variability, and speech context, revealing a causal involvement of left superior temporal cortex in talker adaptation during connected speech. However, it is important to note the large number of degrees of freedom that are available in the design and implementation of brain stimulation studies, including details of the behavioral paradigm, as well as the location, magnitude, and polarity of electrical stimulation. Consequently, future work remains to both replicate and extend the observations from this study.

Our behavioral paradigm involved manipulations that affect speech processing efficiency (Choi, Hu, & Perrachione, 2018; Choi & Perrachione, 2019). However, talker adaptation affects not only speech processing efficiency, but also the phonological and lexical decision outcomes of speech perception (Johnson, 1990; Kleinschmidt & Jaeger, 2015; Francis et al., 2006; Laing et al., 2012). Similarly, talker variability during encoding has differential effects on short-term long-term memories for speech (Lim, Shinn-Cunningham, & Perrachione, in press; Palmeri, Goldinger, & Pisoni, 1993). Future work is therefore needed to understand how left superior temporal lobe is causally involved in recalculating acoustic-phonemic correspondences associated with talker adaptation, and how its role in talker-adaptation processes affects short- and long-term memories for speech.

We found similar behavioral effects of anodal and cathodal stimulation on speech processing efficiency, but hypothesized that the mechanistic bases for these disruptions were nonetheless differentiable. While the present study used a between-subjects design to parsimoniously establish the efficacy of HD-tDCS in studying talker adaptation, this design choice nonetheless precluded the ability to compare the relative effects of anodal vs. cathodal stimulation within individual participants. Future studies may be able to gain better mechanistic insight into how and why these polarities differentially disrupt talker adaptation by comparing effect sizes under a within-subjects design.

Finally, although we found that HD-tDCS of left superior temporal cortex induced a significant and context-specific disruption of talker adaptation, this does not preclude the possibility that other areas of the brain are also causally involved in speech adaptation, or that this region also participates in speech adaptation on other timescales. Stimulation of other sites implicated in talker adaptation (especially the right superior temporal lobe (Zhang et al., 2016; Belin & Zatorre, 2003; Perrachione et al., 2016) and superior parietal lobe (Wong, Nusbaum, & Small, 2004)) must be undertaken in future studies. Similarly, these results should be validated by stimulation at other intensities and using other stimulation paradigms (e.g., transcranial alternating current stimulation) or technologies (e.g., transcranial magnetic stimulation) to replicate and extend our observation of a causal, context-specific role for left superior temporal cortex in talker adaptation.

5. Conclusions

The results from this study show that noninvasive neurostimulation of left superior temporal lobe interferes with the usage of local phonetic context to adapt to a talker and enhance speech processing efficiency, demonstrating that this region is causally involved in rapid talker adaptation.

6. Acknowledgments

We thank Elly Hu and Sara Dougherty for their assistance. This research was supported by the NIDCD of the National Institutes of Health under award number R03DC014045 and by a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation to TP.

7. Open-Source Dataset

The stimuli, paradigm scripts, data, and data analysis scripts associated with this study are available from our institutional archive: <https://open.bu.edu/handle/2144/16460>

References

- Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, *71*, 975–989.
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, *14*(16), 2105–2109.
- Bestmann, S., de Berker, A.O., Bonaiuto, J. (2015). Understanding the behavioral consequences of noninvasive brain stimulation. *Trends in Cognitive Sciences*, *19*, 13-20.
- Choi, J. Y., & Perrachione, T. K. (2019). Time and information in perceptual adaptation to speech. *Cognition*. <https://doi.org/10.1016/j.cognition.2019.05.019>
- Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, & Psychophysics*, *80*(3), 784–797.
- 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 Stimulation*, *2*(4), 201–207.
- Dayan, E., Censor, N., Buch, E.R., Sandrini, M. & Cohen, L.G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. *Nature Neuroscience*, *16*, 838-844.
- Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America*, *119*(3), 1712-1726.
- Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, *6*(11), 1216–1223.
- Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, *450*(7168), 425.
- Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, *59*, 675–692.
- Herrmann, B., Henry, M. J., Fromboluti, E. K., McAuley, J. D., & Obleser, J. (2015). Statistical context shapes stimulus-specific adaptation in human auditory cortex. *Journal of Neurophysiology*, *113*(7), 2582–2591.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.
- Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Rajj, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neuroscience*, *30*(12), 653–661.
- Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Experimental Brain Research*, *216*(1), 1-10.
- Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, *88*(2), 642–654.
- Joos, M. (1948). Acoustic phonetics. *Language Monographs*, *23*, 136.

- 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*(1), 161–172.
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.
- Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C., (1999). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *52*(3), 3–6.
- 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 Stimulation*, *6*(4), 644–648.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*(1), 98–104.
- Laing, E. J., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*, *3*, 203.
- Lieenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*(10), 1621–1631.
- Lim, S.-J., Shinn-Cunningham, B. G., & Perrachione, T. K. (2019). Effects of talker continuity and speech rate on auditory working memory. *Attention, Perception, & Psychophysics*, doi: 10.3758/s13414-019-01684-w
- Magnuson, J. S. & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409.
- Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379–390.
- Myers, E. B. 2007. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, *45*(7), 1463–1473.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, *85*, 2088–2113.
- Nitsche, M.A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *Journal of Physiology*, *527*(3), 633–639.
- 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* (pp. 113–134). Tokyo: Ohmsha Publishing.
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex*, *17*(10), 2251–2257.
- Ohn, S. H., Park, C. I., Yoo, W. K., Ko, M. H., Choi, K. P., Kim, G. M., Lee, Y. T., & Kim, Y. H. (2008). Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport*, *19*(1), 43–47.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.
- Peelle J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Frontiers in Human Neuroscience* *6*:309. doi: 10.3389/fnhum.2012.00309
- 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*(6), 1383–1397.
- Roe, J. M., Nesheim, M., Mathiesen, N. C., Moberget, T., Alnæs, D., & Sneve, M. H. (2016). The effects of tDCS upon sustained visual attention are dependent on cognitive load. *Neuropsychologia*, *80*, 1–8.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics*, *73*(4), 1195–1215.

- Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research*, 18(2), 162–171.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 60, 213–224.
- Thair, H., Holloway, A. L., Newport, R., & Smith, A. D. (2017). Transcranial direct current stimulation (tDCS): a beginner's guide for design and implementation. *Frontiers in Neuroscience*, 11, 641.
- Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG Study. *Journal of Neuroscience*, 31(25), 9118–9123.
- 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(1), 48–55.
- Wernicke, C. Der aphasische Symptomenkomplex. Breslau: Cohn & Weigert; 1874.
- Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, 16(7), 1173–1184.
- Zachle, T., Beretta, M., Jäncke, L., Hermann, C.S., & Sandmann, P. (2011). Excitability changes induced in the human auditory cortex by transcranial direct current stimulation: direct electrophysiological evidence. *Experimental Brain Research*, 215, 135-140.
- Zatorre R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946–953.
- Zhang, C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., ... Wang, W. S.-Y. (2016). Functionally integrated neural processing of linguistic and talker information: An event-related fMRI and ERP study. *NeuroImage*, 124(Pt A), 536–549.
- Zoefel, B., & Davis, M. H. (2017). Transcranial electric stimulation for the investigation of speech perception and comprehension. *Language, Cognition and Neuroscience*, 32(7), 910-923.