Stimulation of right temporal cortex enhances talker typicality judgments

Sahil Luthra[1], Hannah Mechtenberg[2], Hannah E. Olson[2], Holly Zaharchuk[2] & Emily B. Myers[2]
[1]Stony Brook University, [2]University of Connecticut

Corresponding Author:
Sahil Luthra
Department of Psychology
Stony Brook University
100 Nicolls Rd., Stony Brook, NY 11790
sahil.luthra@stonybrook.edu

Abstract

Listeners are exquisitely sensitive to the ways that different talkers produce speech sounds, and such sensitivity may allow listeners to condition speech perception on talker information. A growing body of neuroimaging data suggests that while phonetic processing is largely supported by the brain's left hemisphere, the integration of talker information during speech perception recruits the right hemisphere, especially the right temporal cortex. However, an open question is whether recruitment of the right hemisphere is strictly necessary for integrating talker information and phonetic detail or simply incidental. Here, we leverage transcranial magnetic stimulation (TMS) to test for a causal role of the right middle temporal gyrus (RMTG) in accessing a listener's beliefs about a talker's phonetic signature. During an initial exposure phase, participants heard pairs of talkers, one of whom produced voiceless stop consonants (e.g., /k/) with a relatively short voice onset time (VOT) and one who employed relatively long VOTs; listeners learned what variation was typical of each talker without disruption from TMS. During a subsequent test, TMS was administered to the RMTG, LMTG, or a control site (scalp vertex) before listeners judged whether VOT variants were typical of a learned talker. We found that listeners were best able to determine what phonetic variation was typical of a talker following RMTG stimulation; stimulation of the LMTG did not influence performance relative to vertex stimulation. Overall, results indicate that the balance of activation across the two hemispheres is critical for accessing talker-specific phonetic detail.

Keywords:
Perception: Speech
Hemispheric specialization
Transcranial magnetic stimulation
Language: Phonology

Introduction

Cognitive neuroscientists have long appreciated the left hemisphere's role in speech perception, as established by neuropsychological (e.g., Geschwind, 1970; Wada & Rasmussen, 1960; Wernicke, 1874) and functional neuroimaging data (for review, see Price, 2012). While frank impairments in speech perception are not typically observed following right hemisphere stroke (Turkeltaub & Branch Coslett, 2010), portions of the right posterior temporal cortex have nevertheless been implicated in the early analysis of speech sounds (Hickok & Poeppel, 2007; Hickok et al., 2008; Poeppel, 2001).

Strikingly, right hemisphere damage can result in other communication deficits, including impairments in processing vocal identity (Van Lancker & Canter, 1982). Functional neuroimaging data consistently implicate the right hemisphere in voice processing (Maguinness et al., 2018; von Kriegstein, et al., 2010), with these regions exhibiting sensitivity to factors including voice familiarity (Bethmann et al., 2012), how difficult it is to categorize a voice into socially relevant categories (e.g., gender categories; Andics et al., 2013), and whether an utterance is semantically consistent with a talker's identity (Tesink et al., 2009). Accordingly, the right hemisphere may support the integration of talker information during language comprehension, though the specific functional contributions of different regions remain underspecified (Luthra, 2021).

Even within a dialect, individual talkers may differ substantially in the acoustic-phonetic realization of speech sounds. Accordingly, listeners may benefit from incorporating information about who is talking to determine what is being said. The integration of talker detail might be achieved computationally through conditioning the identity of a speech sound on talker information (e.g., their gender; for a Bayesian formalism, see Kleinschmidt, 2019). Consistent with this proposal, neural data suggest both talker information and phonetic detail are encoded in the right temporal cortex, and interactions between the bilateral temporal cortices support listeners in leveraging talker-specific phonetic information during speech processing (Luthra et al., 2020; Luthra et al., 2023a; Myers & Mesite, 2014).

Of particular relevance to the current work, functional neuroimaging data indicate that the right hemisphere is sensitive to whether a phonetic variant is typical of a talker. In one such study (Myers & Theodore, 2017), listeners were exposed to two talkers who differed in their use of voice onset time (VOT), a primary cue for differentiating stop consonants (Lisker & Abramson, 1964); both talkers produced voiced stops (e.g., /g/) with a characteristic short VOT, but differed in whether they produced voiceless stops (e.g., /k/) with a typical or exaggeratedly long VOT. After participants learned the phonetic signatures of each talker, they heard both talker-typical and talker-atypical variants in an MRI scanner and performed a phonetic categorization task. Notably, a portion of the right middle temporal gyrus (RMTG) extending into the right supramarginal gyrus (RSMG) exhibited sensitivity to whether the phonetic variant was consistent with the talker's typical pronunciation.

Recently, our team conducted a transcranial magnetic stimulation (TMS) study to test for a causal role of this RMTG region in talker-specific phonetic processing (Luthra et al., 2023b).

Specifically, we assessed whether RMTG activation is necessary for listeners to update their beliefs about how different talkers produce their speech sounds (Kleinschmidt & Jaeger, 2015). We applied TMS to the LMTG, RMTG, and a control site (scalp vertex) while participants listened to pairs of talkers who differed in their use of VOT for voiceless stops. We found that TMS applied during a training phase did not modulate listeners' ability to learn what phonetic variation was typical of each talker, as assessed in test blocks.

While numerous factors could explain this null effect, one possibility is that the RMTG may not be necessary for *encoding* the link between talker and phonetic detail, but instead for *retrieval* of this information. In the Myers and Theodore (2017) study, the crucial RMTG effect was observed in a post-learning session; online learning was not measured in the scanner.

The current TMS study is a follow-up to our previous work, wherein we evaluate not whether the RMTG is causally involved in learning a talker's phonetic signature but rather in accessing this signature at test. We investigated this by allowing learning to proceed unimpeded and assessing whether TMS applied during test blocks influenced the ability to determine what phonetic variation was typical of a talker.

Materials & Methods

Stimuli

Stimuli consisted of three minimal pair continua (*bowl*-*pole*, *dime*-*time*, and *gain*-*cane*); the continua were produced by three female talkers (dubbed Carol, Joanne, and Sheila, respectively) and three male talkers (Alvin, Don, and Peter, respectively). These stimuli were taken from our previous TMS study (Luthra et al., 2023b). In brief, stimuli were constructed as follows. Voiced endpoints for each continuum (*bowl*, *dime*, or *gain*) were synthesized using LPC based on natural recordings from different female native speakers of American English. We used different places of articulation and different vowels to discourage generalization of learning across talkers. Continuum steps were created by successively increasing VOT (specifically, by iteratively turning the first voiced frame into a voiceless one; see Allen & Miller, 2004). Stimuli were approximately equal in duration across continua; *bowl*-*pole* stimuli and *gain*-*cane* stimuli (both originally created for Theodore and Miller, 2010) were all 599 ms, whereas *dime*-*time* stimuli (constructed for Allen and Miller, 2004) were 546 ms.

To increase the perceptual dissimilarity of the female voices, we shifted the pitch contours of two of the female talkers in Praat (Boersma & Weenik, 2017); this resulted in three talkers with distinct fundamental frequencies (Carol: 191 Hz, Joanne: 176 Hz, Sheila: 259 Hz). Finally, for each set of female stimuli we constructed associated male continua by modifying the median pitch and the formant spacing (formant shift ratios ranging from 0.70 - 0.85) using the Praat Vocal Toolkit (Corretge, 2026), resulting in three male talkers with distinct fundamental frequencies (Alvin: 113 Hz, Don: 93 Hz, Peter: 125 Hz). We opted to include one male talker and one female talker in each pair to ensure that the talkers were maximally dissimilar and to encourage talker-specific learning. For each stimulus, we presented two amplitude variants (one with RMS amplitude set to 0.070 Pa, one with RMS amplitude set to RMS amplitude was set to 0.055 Pa).

The VOT values presented for each continuum are summarized in Table 1. We employed longer VOTs for voiceless stops with more posterior places of articulation, consistent with natural productions of these consonants (Lisker & Ambramson, 1964). Additionally, we presented slightly different VOT values during training and test. This ensured that participants were not learning a specific exemplar for each talker (e.g., that Sheila always produces *cane* with a VOT of 175 ms) but rather a general rule (e.g., that Sheila produces *cane* with a relatively long VOT and Peter produces it with a relatively short one). While the long-VOT voiceless stimuli are less typical than their short-VOT counterparts, previous work indicates that talkers do produce VOTs in this range when producing words in isolation and that listeners reliably map these exaggerated VOTs to the intended phoneme category (e.g., Allen & Miller, 2004).

Talkers	**Continuum**	**Voice-Onset-Time (ms)**		
		Voiced	*Voiceless (Short)*	*Voiceless (Long)*
Alvin/Carol	bowl-pole	20	Training: 60, 70 Test: 65	Training: 150, 160 Test: 155
Don/Joanne	dime-time	15	Training: 70, 80 Test: 75	Training: 160, 170 Test: 165
Peter/Sheila	gain-cane	20	Training: 80, 90 Test: 85	Training: 170, 180 Test: 175

Table 1. Voice-onset-time (VOT) values (in milliseconds) for the stimuli used in this study. Slightly different VOT values were presented during the training and test phases.

Participants

Forty participants were recruited for the current study from the University of Connecticut community. Four participants were excluded for: being a native speaker of a language other than English, taking a contraindicated medication prior to the TMS session, inability for the research team to find the resting motor threshold, and participant withdrawal due to scheduling conflicts. The final sample therefore included 36 participants (Age: 18-40, *M* = 23.72, *SD* = 5.84; Sex: 20 female, 16 male; Race: 27 White, 4 Black, 4 Asian, 1 more than one race; Ethnicity: 33 not Hispanic or Latino, 3 Hispanic or Latino). All were right-handed, spoke only North American English prior to age 13, reported normal or corrected-to-normal vision, and reported no hearing deficits, language disorders, or neurological impairment. All participants gave informed consent according to the guidelines of the University of Connecticut Institutional Review Board (IRB protocol H21-0046) and received monetary compensation. Our choice of sample size was guided by past studies demonstrating that listeners can learn what phonetic variation is typical of individual talkers (Allen & Miller, 2004; Myers & Theodore, 2017; Theodore & Miller, 2010; mean *N* = 19) and by previous work showing influences of TMS on speech perception (Bestelmeyer et al., 2011; Heimrath et al., 2019; Kennedy-Higgins et al., 2020; Meyer et al., 2018; Nixon et al., 2004; Romero et al., 2006; Smalle et al., 2015; mean *N* = 18). We opted for a sample of 36 to ensure that our power exceeded that of similar studies in this literature.

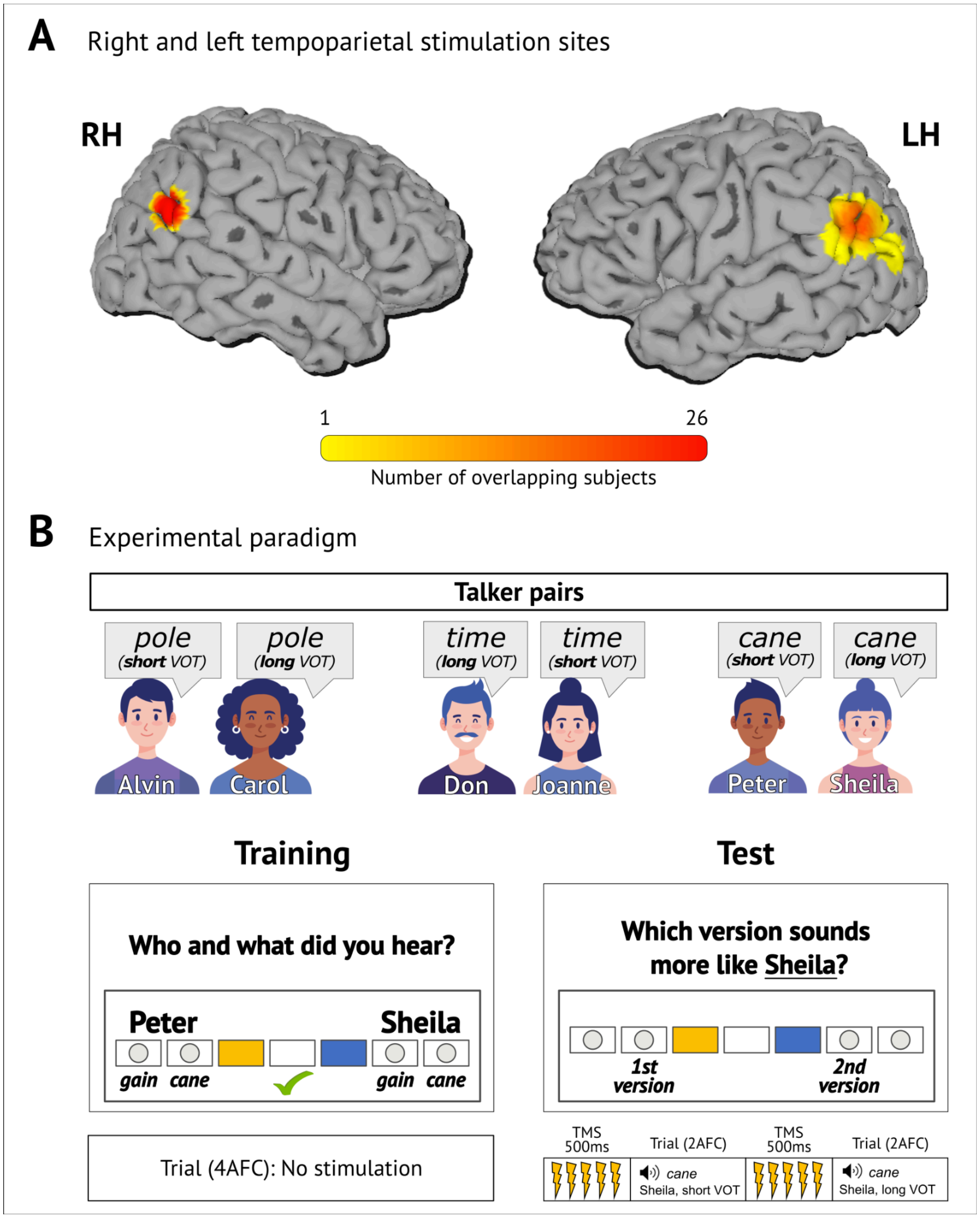

Figure 1. (A) Stimulation sites in the RMTG and LMTG. The RMTG site was localized based on a functional region from Myers & Theodore (2017). The LMTG was the left homologue of that region. (B) Experimental paradigm and depiction of the talker pairs and the associated minimal pairs and VOT lengths; the specific assignment of VOT to talker (e.g., whether Joanne produced *time* with long or short VOT) was counterbalanced across participants. For each stimulation site (e.g., RMTG, LMTG, vertex), participants heard a new talker pair and completed training and test blocks. The training block consisted of a 4AFC decision indicating who was talking (e.g., Peter or Sheila) and what they said (e.g., *gain* or *cane*). Feedback was given after each trial, and there was no stimulation. At test, stimulation was given prior to each trial. Participants only heard the female talker and indicated whether the short or long VOT version was more typical of that talker in a 2AFC. No feedback was given.

Procedure

The current study occurred over two sessions completed on two different days. During session one, after providing consent, all participants completed a suite of questionnaires including questions about demographics, language history and background, handedness, sleep patterns, and neurological history. Hearing thresholds were also evaluated using pure-tone audiometry from 500 Hz to 4000 Hz in 5dB steps. All participants had an average pure-tone threshold of 20 dB or better, indicating no more than minor hearing loss (mean threshold in better ear = 1.08dB, range = -5–17.5). Participants were screened for the presence of possible language and cognitive deficits. A standardized language battery (Fidler et al., 2011) was conducted; all participants scored below 1 on the language battery, indicating no presence of developmental language disorder or only minimal language deficits in childhood (Fidler et al., 2011). Cognitive screening included the Cognitive Linguistic Quick Test-Plus (CLQT+, Helm-Estabrooks, 2001); all participants received a composite score of 4 on the CLQT+, suggesting cognitive functioning within normal limits for their age. Finally, each participant also completed the Quick Speech in Noise (QuickSIN) task (Killion et al., 2004); because of limited variability on this task, it was not used further in analyses.

During session one, we acquired a T1-weighted image of each participant's brain anatomy using MRI at the Brain Imaging Research Core at the University of Connecticut. The images were acquired on a Siemens 3T Prisma scanner with a 64-channel headcoil using a T1-weighted magnetisation-prepared rapid acquisition gradient echo (MP-RAGE) sequence (TR = 2300 ms, TE = 2.98 ms, FOV = 238 mm, flip angle = 9 degrees, 1 mm sagittal slices). The anatomical images were loaded into the Localite Navigator system (Localite, St. Augustin, Germany) to accurately position the TMS coil relative to each participant's anatomy during stimulation. While participants were in the scanner, we also collected two short functional sequences (~10 minutes) used in a different study. During the scans, participants wore OptoACTIVE noise-canceling headphones (OptoACTIVE, Optoacoustics Ltd, Moshav Mazor, Israel) to provide hearing protection and to deliver audio.

Participants completed the experimental task and all TMS stimulation during the second session. Upon beginning the second session, participants were evaluated for TMS safety and screened for any contraindicated medications. Participants were then brought to the TMS room and were prepped for stimulation. Participant head position relative to the TMS coils was tracked using the Localite navigation system (Localite GmbH, Bonn, Germany). We determined each participant's level of stimulation by finding the resting motor threshold (RMT) of the right pollicis brevis muscle (muscle on the palm below the thumb). The RMT is the minimum amount of stimulation required to produce a reliable (e.g., 3 times out of 5 stimulations) 10 μV amplitude muscle contraction during a relaxed state. To assist in finding this threshold, an approximate location in the left motor region of the brain was determined based on each participant's anatomical scan (Yousry et al., 1997). The motor-evoked potentials were recorded using electrodes placed on the participant's hand which fed into a Biopac MP160 system (Biopac Systems Inc., Goleta, CA). All stimulation was delivered using a MagPro X100 TMS device. During motor thresholding, an uncooled butterfly C-B60 coil was used, while during the task, stimulation was delivered with a cooled butterfly coil in active/sham configuration (Cool B-65

A/P; MagVenture, Inc., Atlanta, GA). Uncooled coils are lighter in weight and therefore more maneuverable, making them preferable for precise motor thresholding, but may elicit slightly lower motor thresholds compared to cooled coils. Indeed, motor thresholds evoked with the uncooled C-B60 coil have been found to be approximately 1.5% of the maximum stimulator output (MSO) lower than those evoked by the cooled B-65 (Wang et al., 2023). Accordingly, our protocol may have involved a slight underdosing of stimulation during the experimental task.

Once the right thumb region was found, the appropriate stimulation threshold was determined using the Motor Threshold Assessment Tool (Awiszus & Borckardt, 2011), which includes the adaptive Parameter Estimation by Sequential Testing (PEST) function for determining motor thresholds through an adaptive stairstepping procedure. Ninety percent of RMT was used during task stimulation, resulting in an average of 44.5% MSO with a range of 27%-61%.

The experimental task is schematized in Figure 1B and is the same as was used in the Luthra et al. (2023b) study. There were three total blocks, each with a unique male/female talker pair: Alvin/Carol, Don/Joanne, Peter/Sheila. At the start of each block, participants were familiarized with the productions of each talker (e.g., Alvin and Carol each saying *bowl* and *pole*). Then, during training, participants made a 4-alternative forced choice (4AFC) that indicated who was talking (e.g., Alvin or Carol) and what they said (e.g., *bowl* or *pole*). After each response, participants were given feedback. No stimulation was administered during the training phase. At test, participants only heard productions from the female talker and made a 2-alternative forced choice (2AFC) to indicate whether the first or second token was more typical of that talker. No feedback was provided. To discourage generalization across talkers, we alternated the characteristic VOT of same-sex talkers across blocks; thus, if a listener heard a female talker with long-VOT voiceless productions in block 1, they would hear a female talker with short-VOT voiceless productions in block 2, and a female talker with long-VOT voiceless productions in block 3, and vice versa.

TMS was administered in between trials during the test portion of the task. The stimulation protocol adhered to current safety standards (Rossi et al., 2009, 2021). Immediately preceding each test trial, five biphasic bursts (pulses) of 10 Hz frequency stimulation at 90% of RMT were delivered to one of three sites: RMTG, LMTG, or the control site (vertex). Stimulation at 10 Hz has previously been linked to inhibiting function in the superior temporal gyrus (e.g., Bestelmeyer et al., 2011; Kennedy-Higgins et al., 2020). Accordingly, we hypothesized that stimulation of the RMTG would lead to worse behavioral performance on our task (i.e., decreased ability to determine if a phonetic variant was typical of a talker); however, effects of TMS stimulation can depend on task and stimulation site, not just stimulation frequency (e.g., Hartwigsen & Silvanto, 2023), a point we return to in the Discussion. We opted to use the scalp vertex as a control site to account for the physical stimulation (audible clicks, tapping sensation) of receiving TMS; however, we also provide a comparison of the present results to data from a separate group of participants who received no stimulation. Stimulation of scalp vertex may influence activation of the Default Mode Network (Jung et al., 2016) but is not known to influence networks involved in processing speech.

The RMTG stimulation site was functionally defined based on a region previously shown to be sensitive to whether a phonetic production was typical of a talker (Myers & Theodore, 2017). Notably, other right hemisphere regions may also support talker-specific phonetic processing (e.g., the right superior temporal sulcus; Luthra et al., 2023a); accordingly, additional work is needed to more precisely characterize the functional neuroanatomy of the right hemisphere in speech perception. For each participant, our RMTG cluster (Talairach coordinates: x = 55, y = -60, z = 17) was resampled from Talairach & Tournoux space onto each participant's anatomy using *3dFractionize* in AFNI (Cox, 1996). The LMTG site (Talairach coordinates: x = -52, y = -61, z = 23) was selected as an anatomical homologue to the RMTG site, and the vertex was selected such that the coil was perpendicular to the central fissure. To visualize the overlap in stimulation sites across participants, we drew an 8mm sphere around each participant's stimulation coordinates (X, Y, Z) and resampled into Talairach & Tournoux space (Figure 1A).

The order of the stimulation sites was counterbalanced in three combinations (LMTG-Vertex-RMTG, RMTG-LMTG-Vertex, Vertex-RMTG-LMTG), as was the order of the talker pairs (Alvin/Carol-Don/Joanne-Peter/Sheila, Don/Joanne-Peter/Sheila-Alvin/Carol, Peter/Sheila-Alvin/Carol-Don/Joanne), such that each unique combination of talker pair and stimulation site was presented at every ordinal position across participants. In addition, the assignment of VOTs to talkers was counterbalanced across participants as described in the previous paragraph (Long-Short-Long, Short-Long-Short). In total, there were 18 counterbalancing orders (3 talker pair orders x 3 stimulation site orders x 2 typical VOT orders). Note that due to the location of the LMTG and RMTG stimulation sites, our protocol sometimes resulted in the stimulation of the temporalis muscle (which attaches along the lateral plane of the skull), leading to mild jaw twitching. We documented the laterality of the twitches and verified the comfort of participants throughout stimulation. The majority of participants did not report muscular twitching (*N* = 19); of the participants who reported muscular twitching (*N* = 17), seven reported the sensation with LMTG stimulation, four with RMTG stimulation, five at both lateral stimulation sites, and one was unsure about the laterality.

The audio for the experimental task was delivered via a Focusrite Scarlett 2i2 digital audio interface (High Wycombe, England) through foam-wrapped ER-3C insert headphones (Etymotic Research, Elk Grove Village, IL). As the TMS coil discharge can create a loud popping noise, these headphones protected participant hearing while also allowing them to hear the stimuli; experimenters wore earplugs to protect their hearing as well. We used OpenSesame to deliver the task (Mathôt et al., 2012) and participants responded using a button box.

Analyses of the test task were implemented in R (R Core Team, 2025). Summary statistics were calculated using the *summarySE* function in the *Rmisc* package (Hope, 2022). We first conducted an omnibus analysis of the binary response data (long VOT token = 1, short VOT token = 0) using the *mixed* function in the *afex* package (Singmann et al., 2018), which serves as a wrapper to the *lme4* package (Bates et al., 2015) and evaluates the significance of each fixed effect with likelihood ratio tests that consider the improvement in model fit for a full model compared to a reduced model without that effect. Our omnibus model included sum-coded fixed factors for Typical VOT (long VOT talker = 1, short VOT talker = -1) and Stimulation Site (CC1:

Vertex = 1, LMTG = 0, RMTG = -1; CC2: Vertex = 0, LMTG = 1, RMTG = -1) as well as their interaction; we also included by-subject random intercepts and specified a binomial family with a logit link. This approach yields an ANOVA-like output, which is useful for characterizing main effects of and/or interactions with factors that have more than two levels (as was the case for Stimulation Site in the current work). The design of our experiment precluded the inclusion of random by-subject slopes for Typical VOT or Stimulation Site (Barr et al., 2013), since each participant only received one level of Typical VOT for each Stimulation Site; that is, if a participant heard a long-VOT talker while receiving RMTG stimulation, they would not have also heard a short-VOT talker for that stimulation site. Thus, it is not possible to estimate a random slope for Stimulation Site that is not also influenced by the Typical VOT encountered in that block, and vice versa. We note that this aspect of the design may have also driven the lack of Typical VOT x Stimulation Site effects observed in our previous TMS study (Luthra et al., 2023b) and that this point should be considered if null interactions are observed in the current work.

To further characterize any effects of stimulation site on the test task, we conducted two follow-up regression analyses using the *glmer* function in the *lme4* library. These analyses were essentially identical to the previous one, but used treatment coding for the Stimulation Site factor, which allowed one level to be set as a reference level. In the first follow-up analysis, we set the Vertex site as our reference level to compare stimulation at each of the temporal sites (RMTG and LMTG) to the control site (CC1: Vertex = 0, LMTG = 1, RMTG = 0; CC2: Vertex = 0, LMTG = 0, RMTG = 1). Effect sizes were quantified as the difference in the proportion of trials in which listeners selected the long-VOT variant across conditions. The second follow-up analysis used RMTG as the reference level in order to compare RMTG to LMTG stimulation. We did not conduct any statistical analyses of the training data; since TMS was not administered during the training phase, we did not expect any differences in performance between blocks.

Finally, at the suggestion of a reviewer, we compared the control condition in this experiment (Vertex stimulation) to a control group that did not receive any stimulation (Experiment 1, Luthra et al., 2023b). We analyzed the binary response data as described above using the *mixed* function from the *lme4* package with a binomial family function (logit link). We included Typical VOT, Stimulation (Vertex = 1, None = -1), and their interaction as fixed effects and subjects as random intercepts. Note that Typical VOT is a within-subjects manipulation for the control group without stimulation, while it is a between-subjects manipulation for the Vertex stimulation condition (since each participant only heard one test talker at each Stimulation Site). While both samples were drawn from the same population of university students, we note that there may be important cohort differences and present these data as supportive rather than central evidence.

Results

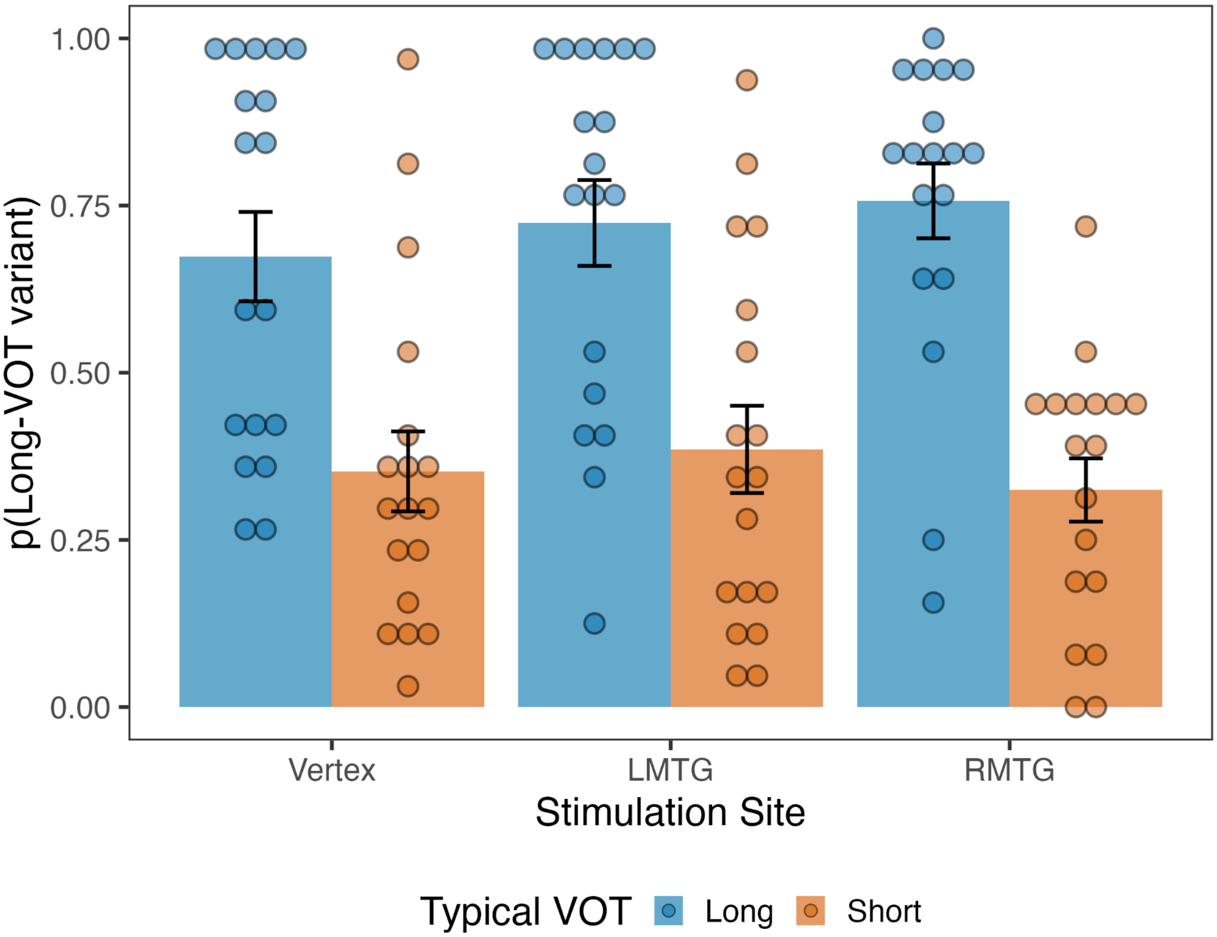

Figure 2. Performance during the test phase, as a function of stimulation site and whether the talker historically had produced their voiceless consonants with long (blue) or short (orange) VOT. Talker-specific phonetic processing is indexed by the difference in how often listeners selected the long-VOT variant as more typical of a talker across conditions. We observed the largest typicality effect when participants received RMTG stimulation (effect size = 0.43), with equivalent effects following LMTG (effect size = 0.34) and vertex (effect size = 0.32) stimulation. Circles indicate data for individual participants. Error bars indicate standard error.

Participants exhibited ceiling levels of accuracy on the training task, with 99.5% accuracy in identifying which of the two talkers they heard (*SE* = 0.1%, 95% CI [99.3%, 99.7%]) and 97.4% accuracy in identifying which of the two words (e.g., *gain* or *cane*) they heard (*SE* = 0.5%, 95% CI [96.5%, 98.3%]). These results indicate that our stimuli were perceptually distinct and easily identifiable.

During test blocks, participants made judgments on which of two phonetic variants was typical of an individual voice. Note that the critical effect is the difference in the proportion of long-VOT responses as a function of whether the talker had historically used a long or short VOT for voiceless consonants; we refer to the difference in proportions as the *talker typicality effect*. Of interest is whether this effect differs as a function of stimulation; accordingly, we are principally interested in an interaction of Stimulation Site and Typical VOT. Visual inspection of results (Figure 2 and Table 2) suggests that regardless of stimulation site, listeners were more likely to select the long-VOT variant when the talker had previously used long-VOT variants. However, this talker typicality effect appears to be more pronounced following RMTG stimulation compared to LMTG or Vertex stimulation.

Statistical analysis of the data supports these observations. Our omnibus model yielded a significant main effect of Typical VOT, $\chi^2(1) = 496.15$, *p* < .001, indicating that listeners were able to evaluate which phonetic variant was typical of each talker (Long: *M* = 0.75, *SE* = 0.04, 95% CI [0.66, 0.82]; Short: *M* = 0.36, *SE* = 0.03, 95% CI [0.29, 0.43]). There was also a significant Stimulation Site x Typical VOT interaction, $\chi^2(2) = 30.67$, *p* < .001, suggesting that the ability to identify which variant was talker-typical differed as a function of stimulation site. This interaction effect is depicted in Figure 2; summary statistics are provided in Table 2. We also observed a marginal effect of Stimulation Site, $\chi^2(2) = 5.58$, *p* = .061, reflecting a slightly lower likelihood of selecting the long-VOT variant after Vertex stimulation (*M* = 0.51, *SE* = 0.05, 95% CI [0.40, 0.62]) compared to RMTG (*M* = 0.54, *SE* = 0.05, 95% CI [0.44, 0.64]) or LMTG (*M* = 0.55, *SE* = 0.05, 95% CI [0.44, 0.66]) stimulation.

To characterize the significant interaction between Stimulation Site and Typical VOT, we conducted a follow-up analysis that compared the typicality effect in the two MTG sites to the typicality effect in the control (Vertex) condition. While we had hypothesized that this effect would be diminished following RMTG stimulation, this follow-up analysis indicated that the ability to determine whether a variant was typical of a talker was significantly enhanced following RMTG stimulation compared to following Vertex stimulation, β = 0.49, *SE* = 0.12, 95% CI [0.25, 0.73], *z* = 4.01, *p* < .001. The typicality effect following LMTG stimulation was not significantly different from the typicality effect in the Vertex condition, β = -0.17, *SE* = 0.13, 95% CI [-0.42, 0.08], *z* = -1.36, *p* = .175. A second follow-up analysis with RMTG as the reference level also showed that RMTG stimulation drove larger talker typicality effects compared to LMTG stimulation, β = -0.66, *SE* = 0.13, 95% CI [-0.91, -0.41], *z* = -5.24, *p* < .001.

When comparing Vertex stimulation to no stimulation in a separate group of participants, we observed a main effect of Stimulation, $\chi^2(1) = 5.38$, *p* = .020, reflecting an overall higher proportion of long-VOT responses with Vertex stimulation (*M* = 0.51, *SE* = 0.05, 95% CI [0.40, 0.62]) compared to no stimulation (*M* = 0.35, *SE* = 0.04, 95% CI [0.27, 0.43]). We also observed a main effect of Typical VOT, $\chi^2(1) = 36.49$, *p* < .001, with listeners judging the typicality of a phonetic variant in line with their previous experience with that talker (Long: *M* = 0.60, *SE* = 0.05, 95% CI [0.51, 0.69]; Short: *M* = 0.29, *SE* = 0.03, 95% CI [0.22, 0.36]). Most crucially, we did not observe an interaction between Stimulation Status and Typical VOT, $\chi^2(1) = 0.13$, *p* = .720, indicating that Vertex stimulation did not influence the typicality effect relative to a control group. Summary statistics are provided in Table 2.

Stimulation Site	**Typical VOT**						
	Long			*Short*			*Long - Short*
	M	*SE*	*95% CI*	*M*	*SE*	*95% CI*	*Effect size*
Vertex	0.674	0.067	[0.533, 0.815]	0.352	0.060	[0.226, 0.478]	0.322
LMTG	0.724	0.064	[0.589, 0.859]	0.385	0.065	[0.247, 0.523]	0.339
RMTG	0.757	0.056	[0.639, 0.875]	0.325	0.047	[0.225, 0.425]	0.432
No stimulation	0.556	0.062	[0.429, 0.683]	0.245	0.041	[0.162, 0.328]	0.311

Table 2. Means and standard errors for the proportion of long-VOT responses by Stimulation Site (Vertex, LMTG, RMTG) and Typical VOT (Long, Short) in the test phase. The effect size at each stimulation site (calculated as the mean proportion of long-VOT responses for Long VOT talkers vs. Short VOT talkers) is also shown. Data from the control group in Experiment 1 of Luthra et al. (2023b) is provided for comparison (No stimulation).

Discussion

The current study tested whether TMS of the LMTG and RMTG affected a listener's ability to determine if a speech sound was typical of a talker. Listeners learned the association between a talker and their particular phonetic variant without disruption from TMS. Stimulation at test was intended to disrupt the integration of talker information (e.g., Sheila is speaking) with phonetic detail (e.g., this sound has a short VOT). In all stimulation conditions (LMTG, RMTG, and vertex control), listeners showed a significant ability to judge which phonetic variant was most typical of a given talker. However, during the RMTG stimulation block, performance was *enhanced*, yielding a significantly larger effect of talker typicality.

Because speech sounds can be so variable across talkers (Liberman et al., 1967; Newman et al., 2001; Peterson & Barney, 1952), an understanding of how a talker's phonetic categories lie in acoustic space may facilitate speech perception (see Nygaard et al., 1994; Nygaard & Pisoni 1998; Levi et al., 2015). For instance, familiar voices are more easily recognized in noise (Souza et al., 2013), showing an approximately 4 dB signal-to-noise ratio advantage. Listeners are often explicitly aware of even subtle sub-phonetic differences in talkers' speech patterns (Allen & Miller, 2004; Myers & Theodore, 2017; Theodore & Miller, 2010), a finding replicated robustly in our behavioral data.

The present results build on a previous study employing a similar paradigm (Luthra et al., 2023b) wherein we applied TMS to the RMTG as listeners learned what phonetic variation was typical of a talker; in that study, we found that the RMTG did not causally influence listeners' ability to learn the link between talker information and phonetic detail, though we acknowledge that many factors could have contributed to those null results. In the current work, we found that TMS to this same region enhanced a listener's ability to access learned representations of a talker's phonetic signature.

Notably, the specific pattern of results was unexpected. Given the role of the RMTG in integrating phonetic detail with talker identity (Myers & Theodore, 2017) and previous work demonstrating an inhibitory influence of this stimulation paradigm on speech perception (Kennedy-Higgins et al., 2020), we had hypothesized that RMTG stimulation would reduce sensitivity to talker typicality information. However, many factors impact whether TMS facilitates or inhibits cortical responses (see Hartwigsen & Silvanto, 2023 for review), including the baseline excitability of the cortical site and whether this site is excited or suppressed given task demands. Indeed, Bergmann and Hartwigsen (2021) caution against inferring a direct causative action mechanism between stimulation of a site and a behavioral outcome. Accordingly, we see at least three possibilities for interpreting the current findings.

One possibility is that our stimulation schedule had a facilitatory, rather than inhibitory effect on local cortical activity. This would be consistent with prior evidence linking the RMTG to the integration of talker information and phonetic detail (e.g., Myers & Theodore, 2017). While low-frequency stimulation tends to be facilitatory and high-frequency stimulation inhibitory, the stimulation frequency at which behavioral effects flip from facilitatory to inhibitory depends crucially on the excitability of the structure being stimulated, which is itself dictated by task effects (Silvanto & Cattaneo, 2017). In considering the current results, it is notable that hearing talker-atypical phonetic variants leads to deactivation of the RMTG (Myers & Theodore, 2017); this may have influenced whether our stimulation paradigm was facilitatory or inhibitory. To speculate, if processing talker-specific phonetic detail leads to deactivation in the RMTG, TMS to this area may have brought activation levels closer to baseline, thus having a facilitatory rather than inhibitory effect.

A second possibility is that stimulation in the RMTG enhanced contralateral LMTG excitability. For instance, Andoh and Paus (2011) showed that rTMS at 10 Hz increased the hemodynamic response to language stimuli in the hemisphere contralateral to stimulation. These authors interpreted this pattern as reflecting a release from inhibition of the contralateral hemisphere evoked when activity is disrupted via rTMS. Under this logic, stimulation of RMTG site in the current study might have enhanced the excitability of the contralateral temporal lobe, leading to the interpretation that left, not right temporal sites handle integration of talker information and phonetic detail. We view this possibility as the least likely, given the postulated role for the right hemisphere in voice processing (see Luthra, 2021, for review).

Finally, it is possible that TMS to the RMTG altered the distribution of processing across the two hemispheres, whether through local and/or distal effects. Previous studies have demonstrated that when listeners must integrate talker information and phonetic detail, there is increased functional connectivity between left hemisphere regions implicated in phonetic processing and right hemisphere structures linked to voice processing (e.g., Luthra et al., 2023a). Under this view, the enhanced talker-specific phonetic processing observed in the current study results from the reweighting of processing across hemispheres, rather than reflecting a local effect of stimulation. In the absence of concurrent neuroimaging, it is difficult to determine the precise mechanism through which the observed neuromodulation effects arose.

We interpret the current data as demonstrating that disrupting the coordination between the left and right hemispheres (whether through local neuromodulation, contralateral effects, or network-level influences) influences listeners' ability to access talker-specific phonetic detail. The current study therefore motivates future work assessing possible subclinical deficits in talker-specific phonetic processing in individuals with damage to either the left or right hemisphere of the brain.

Our TMS studies emphasize a distinction between learning and accessing mental representations of phonetic distributions, since we did not observe effects of stimulation in our previous study, when TMS was applied during the learning phase (Luthra et al., 2023b), but did in the current work, when TMS was applied at test (when listeners had to access these representations). Notably, similar distinctions have been observed elsewhere in the literature. In the context of non-native phonetic learning, for instance, successfully learning to differentiate novel speech sounds is initially supported by frontal regions, whereas perceptual sensitivity in temporal regions is retuned over longer timescales (e.g., Myers & Swan, 2012; Myers, 2014). In native language phonetic variation, both right frontal and right temporal cortices have been linked to phonetic recalibration (e.g., Myers & Mesite, 2014; Luthra et al., 2020). Considering the present findings in this light, it is possible that right frontal regions may play a stronger role in learning how a talker produces their speech sounds, whereas right temporal regions might be necessary for accessing learned representations. We defer this question to future empirical work.

Our results are broadly consistent with the view that the left and right hemispheres play different functional roles in communication, with the left hemisphere being involved in phonetic processing and the right in talker processing (Van Lancker, 1980). What may drive such a preference? Some researchers have observed that the left and right temporal cortices appear to sample the acoustic signal asymmetrically, with the left hemisphere responding more strongly to information that varies over relatively short time scales (on the order of 25 ms) and the right hemisphere exhibiting a preference for low-frequency modulations (on the order of 200 ms; Poeppel, 2003). This Asymmetric Sampling in Time (AST) hypothesis could, in theory, explain the relative contributions of the two hemispheres, since phonetic distinctions typically involve high-frequency modulations, whereas prosody and voice information are typically conveyed over longer time windows. In the current study, listeners learned how individual talkers differed in their use of VOT, an acoustic-phonetic property that is conveyed over a relatively short timescale. Accordingly, the AST hypothesis would predict that these stimuli should be processed primarily by the left hemisphere. However, the present data indicate that the RMTG is involved in the talker-specific processing of an acoustic cue conveyed over a *short* timescale, a finding also corroborated by past fMRI data (Myers & Theodore, 2017). Overall, then, our data suggest that the division of labor between the left and right hemispheres may be better explained by a functional differentiation between the two hemispheres rather than a physical one (see McGettigan & Scott, 2012), challenging the AST hypothesis.

The current study adds to a line of research aiming to clarify the role of the right hemisphere in the neurobiology of language. Beyond talker-specific processing, the right hemisphere has been shown to be critical for pragmatic aspects of communication (e.g., prosody, emotion, metaphor) as well as social identity processing (Gainotti, 2007). Despite such evidence that the right hemisphere plays a meaningful role in communication, however, communication deficits following right hemisphere damage are typically observed to be mild. We hypothesize that this may be due in part to listeners' ability to capitalize on other (lexical, semantic, syntactic, etc.) contextual cues that might guide perception. This leads to the testable prediction that speech perception in individuals with right hemisphere stroke may be challenged when other sources of context are removed or when demands on talker-specific processing are especially high (e.g., in adapting to an unfamiliar accent). Accordingly, we highlight the importance of future work in clarifying possible subclinical impairments in individuals with right hemisphere damage.

Data Availability Statement: Stimuli, data, and analysis code are available at https://osf.io/z52v8/.

Author Contributions: Sahil Luthra: Conceptualization, Methodology, Data Curation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization **Hannah Mechtenberg:** Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, , Writing - Review & Editing, Visualization **Hannah Olson:** Methodology, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing **Holly Zaharchuk:** Methodology, Investigation, Data Curation, Formal analysis, Writing - Original Draft, , Writing - Review & Editing, Visualization **Emily Myers:** Conceptualization, Supervision, Funding acquisition, Writing - Original Draft, Writing - Review & Editing

Acknowledgments: The authors would like to thank the staff at the Brain Imaging Research Core for providing support throughout data collection and Rachel Theodore for providing the stimuli used in previous instantiations of this paradigm.

Funding Information: Research reported in this publication was supported by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under award number R01 DC013064 (Myers, PI). Sahil Luthra was supported by NIH F32 DC020625. Hannah Mechtenberg was supported on an IBRAiN Fellowship through the Institute of Brain and Cognitive Sciences at the University of Connecticut. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Gender Citation Balance Index (GCBI):
Proportion of categorized DOIs by category: 0.353 MM, 0.176 WM, 0.196 MW, 0.275 WW
GCBI by category: -0.133 MM, -0.425 WM, 0.69 MW, 0.605 WW
Range: <0 = under-cited, 0 = cited at JoCN base rate, >0 = over-cited

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171-3183. https://doi.org/10.1121/1.1701898

Andics, A., McQueen, J. M., & Petersson, K. M. (2013). Mean-based neural coding of voices. *NeuroImage, 79*, 351-360. https://doi.org/10.1016/j.neuroimage.2013.05.002

Andoh, J., & Paus, T. (2011). Combining functional neuroimaging with off-line brain stimulation: Modulation of task-related activity in language areas. *Journal of Cognitive Neuroscience*, *23*(2), 349-361. https://doi.org/10.1162/jocn.2010.21449

Awiszus, F., & Borckardt, J. (2011). TMS motor threshold assessment tool (MTAT 2.0). *Brain Stimulation Laboratory, Medical University of South Carolina, USA*.

Bergmann, T. O., & Hartwigsen, G. (2021). Inferring causality from noninvasive brain stimulation in cognitive neuroscience. *Journal of Cognitive Neuroscience*, *33*(2), 195-225. https://doi.org/10.1162/jocn_a_01591

Bestelmeyer, P. E., Belin, P., & Grosbras, M. H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, *21*(20), R838-R839. https://doi.org/10.1016/j.cub.2011.08.046

Bethmann, A., Scheich, H., & Brechmann, A. (2012). The temporal lobes differentiate between the voices of famous and unknown people: An event-related fMRI study on speaker recognition. *PloS One, 7*(10), e47626. https://doi.org/10.1371/journal.pone.0047626

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenik, D. (2017). Praat: Doing phonetics by computer. https://doi.org/10.1097/aud.0b013e31821473f7

Corretge, Ramon (2026). Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162-173. https://doi.org/10.1006/cbmr.1996.0014

Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, *20*(1), 2-13. https://doi.org/10.1044/1058-0360(2010/09-0096)

Fidler, L. J., Plante, E., & Vance, R. (2013). Erratum. *American Journal of Speech-Language Pathology*, *22*(3), 577. https://doi.org/10.1044/1058-0360(2013/13-0018)

Gainotti, G. (2007). Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: a systematic review. *Neuropsychologia, 45*(8), 1591-1607. https://doi.org/10.1016/j.neuropsychologia.2006.12.013

Geschwind, N. (1970). The organization of language and the brain: Language disorders after brain damage help in elucidating the neural basis of verbal behavior. *Science*, *170*(3961), 940-944. https://doi.org/10.1126/science.170.3961.940

Hartwigsen, G., & Silvanto, J. (2023). Noninvasive brain stimulation: Multiple effects on cognition. *The Neuroscientist*, *29*(5), 639-653. https://doi.org/10.1177/10738584221113806

Heimrath, K., Spröggel, A., Repplinger, S., Heinze, H. J., & Zaehle, T. (2020). Transcranial static magnetic field stimulation over the temporal cortex modulating the right ear advantage in dichotic listening. Neuromodulation: *Technology at the Neural Interface*, *23*(3), 335-340. https://doi.org/10.1111/ner.13023

Helm-Estabrooks, N. (2001). Cognitive Linguistic Quick Test. Pearson Education.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., Barde, L., & Grant, A.. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language*, *107*(3), 179-184. https://doi.org/10.1016/j.bandl.2008.09.006

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402. https://doi.org/10.1038/nrn2113

Hope, R. M. (2022). Rmisc: Ryan Miscellaneous. https://CRAN.R-project.org/package=Rmisc

Jung, J., Bungert, A., Bowtell, R., & Jackson, S. R. (2016). Vertex stimulation as a control site for transcranial magnetic stimulation: a concurrent TMS/fMRI study. *Brain Stimulation*, *9*(1), 58-64. https://doi.org/10.1016/j.brs.2015.09.008

Kennedy-Higgins, D., Devlin, J. T., Nuttall, H. E., & Adank, P. (2020). The causal role of left and right superior temporal gyri in speech perception in noise: A transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *32*(6), 1092-1103. https://doi.org/10.1162/jocn_a_01521

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395-2405. https://doi.org/10.1121/1.1784440

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. https://doi.org/10.1037/a0038695

Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language*, *42*(4), 843-872. https://doi.org/10.1017/s0305000914000506

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431. https://psycnet.apa.org/doi/10.1037/h0020279

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Luthra, S. (2021). The role of the right hemisphere in processing phonetic variability between talkers. *Neurobiology of Language, 2*(1), 138-151. https://doi.org/10.1162/nol_a_00028

Luthra, S., Correia, J. M., Kleinschmidt, D. F., Mesite, L., & Myers, E. B. (2020). Lexical information guides retuning of neural patterns in perceptual learning for speech. *Journal of Cognitive Neuroscience, 32*(10), 2001-2012. https://doi.org/10.1162/jocn_a_01612

Luthra, S., Magnuson, J. S., & Myers, E. B. (2023a). Right posterior temporal cortex supports integration of phonetic and talker information. *Neurobiology of Language, 4*(1), 145-177. https://doi.org/10.1162/nol_a_00091

Luthra, S., Mechtenberg, H., Giorio, C., Theodore, R. M., Magnuson, J. S., & Myers, E. B. (2023b). Using TMS to evaluate a causal role for right posterior temporal cortex in talker-specific phonetic processing. *Brain and Language*, *240*, 105264. https://doi.org/10.1016/j.bandl.2023.105264

Maguinness, C., Roswandowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia, 116*, 179-193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: What's wrong, what's right and what's left? *Trends in Cognitive Sciences*, *16*(5), 269–276. https://doi.org/10.1016/j.tics.2012.04.006

Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in Neuroscience*, *8*, 238. https://doi.org/10.3389/fnins.2014.00238

Meyer, L., Elsner, A., Turker, S., Kuhnke, P., & Hartwigsen, G. (2018). Perturbation of left posterior prefrontal cortex modulates top-down processing in sentence comprehension. *NeuroImage*, *181*, 598-604. https://doi.org/10.1016/j.neuroimage.2018.07.059

Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language, 76*, 80-93. https://doi.org/10.1016/j.jml.2014.06.007

Myers, E. B., & Theodore, R. M. (2017). Voice-sensitive brain networks encode talker-specific phonetic detail. *Brain and Language*, *165*, 33-44. https://doi.org/10.1016/j.bandl.2016.11.001

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, *109*(3), 1181-1196. https://doi.org/10.1121/1.1348009

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The inferior frontal gyrus and phonological processing: an investigation using rTMS. *Journal of Cognitive Neuroscience*, *16*(2), 289-300. https://doi.org/10.1162/089892904322984571

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*(3), 355-376. https://doi.org/10.3758/BF03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42-46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America*, *24*, 175. https://doi.org/10.1121/1.1906875

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, *25*(5), 679-693. https://doi.org/10.1016/S0364-0213(01)00050-7

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication, 41*(1), 245-255. https://doi.org/10.1016/S0167-6393(02)00107-3

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847. https://doi.org/10.1016/j.neuroimage.2012.04.062

R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Romero, L., Walsh, V., & Papagno, C. (2006). The neural correlates of phonological short-term memory: A repetitive transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *18*(7), 1147-1155. https://doi.org/10.1162/jocn.2006.18.7.1147

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008-2039. https://doi.org/10.1016/j.clinph.2009.08.016

Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmöller, J., Carpenter, L.L., Cincotta, M., Chen, R., Daskalakis, J.D., Di Lazzaro, V., Fox, M.D., George, M.S., Gilbert, D., Kimiskidis, V.K., Koch, G., Ilmoniemi, R.J., Lefaucheur, J. P., Leocani, L., Lisanby, S.H., & Hallet, M. (2021). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert Guidelines. *Clinical Neurophysiology*, *132*(1), 269-306. https://doi.org/10.1016/j.clinph.2020.10.003

Silvanto, J., & Cattaneo, Z. (2017). Common framework for “virtual lesion” and state-dependent TMS: The facilitatory/suppressive range model of online TMS effects on behavior. *Brain and Cognition*, *119*, 32-38. https://doi.org/10.1016/j.bandc.2017.09.007

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2018). *afex: Analysis of Factorial Experiments*. *R package version 0.27-2*.

Smalle, E. H., Rogers, J., & Möttönen, R. (2015). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex*, *25*(10), 3690-3698. https://doi.org/10.1093/cercor/bhu218

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, *24*(8), 689-700. https://doi.org/10.3766/jaaa.24.8.6

Tesink, C. M., Petersson, K. M., Van Berkum, J. J., Van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience, 21*(11), 2085-2099. https://doi.org/10.1162/jocn.2008.21161

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *128*(4), 2090-2099. https://doi.org/10.1121/1.3467771

Turkeltaub, P. E., & Branch Coslett, H. (2010). Localization of sub-lexical speech perception components. *Brain and Language, 114*(1), 1-15. https://doi.org/10.1016/j.bandl.2010.03.008

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Research on Language & Social Interaction, 13*(2), 201-277. https://doi.org/10.1080/08351818009370498

Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition, 1*(2), 185-195. https://doi.org/10.1016/0278-2626(82)90016-1

Von Kriegstein, K., Smith, D. R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *Journal of Neuroscience*, *30*(2), 629-638. https://doi.org/10.1523/JNEUROSCI.2742-09.2010

Wada, J., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. *Journal of Neurosurgery*, *17*(2), 266-282. https://doi.org/10.3171/jns.1960.17.2.0266

Wang, Y., Vora, I., Huynh, B. P., Picard-Fraser, M., Daneshzand, M., Nummenmaa, A., & Kimberley, T. J. (2023). Coils are not created equal: Effects on TMS thresholding. *Brain Stimulation*, *17*(1), 1-3. https://doi.org/10.1016/j.brs.2023.11.017

Wernicke, C. (1974). Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis. In *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis* (pp. 1-70). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-65950-8_1

Yousry, T. A., Schmid, U. D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., & Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus: A new landmark. *Brain: A Journal of Neurology*, *120*(1), 141-157. https://doi.org/10.1093/brain/120.1.141