A Causal Role of Human Laryngeal Motor Cortex in Speech Lexical Tone and Voicing Perception

Baishen Liang[1,2], Yanchang Li[1], Wanying Zhao[1], Yi Du[1,2,3,4]*

[1] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China 100101

[2] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China 100049

[3] CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China 200031

[4] Chinese Institute for Brain Research, Beijing, China 102206

* Corresponding author:

Dr. Yi Du

16 Lincui Road, Chaoyang District, Beijing, China 100101

Email: duyi@psych.ac.cn

Abstract

To achieve categorical speech perception, listeners can actively reconstruct motor gestures of the upcoming speech. This is supported by sensorimotor integration function of the dorsal pathway where the “internal models” of articulations implement top-down constraints on auditory speech analyses in the ventral pathway. Moreover, the motor engagement in speech perception is effector-specific, resembling the motor somatotopy during speech production. However, whether and how the laryngeal motor cortex (LMC), which controls the laryngeal muscles that regulate pitch contours and speech voicing, is also involved in speech perception remains largely unknown. To address this question, this study applied transcranial magnetic stimulation (TMS) upon the left or right LMC when participants performed categorical perception tasks on Mandarin lexical tone (pitch contour contrast) or consonant (voice onset time contrast) presented with or without noise. Results showed that, inhibitory stimulation (repetitive TMS or continuous theta-burst stimulation) on the LMC impaired both tone and consonant perception. Furthermore, TMS effects were especially evident when syllables were masked by noise. These results reveal a causal role of the LMC in speech perception that (1) it mirrors its distinctive functions in controlling pitch and voicing during production and (2) it is upregulated and shifts from a left-lateralized to a bilateral engagement in adverse listening conditions as a means of compensation. This study thus provides new evidence to support the motor theory of speech perception, and more generally, active perception and embodied cognition.

Keywords:

laryngeal motor cortex, sensorimotor integration, categorical speech perception, lexical tone, voicing

Introduction

Human speech is characterized by distinctive acoustic-articulatory features that form phonological contrasts (i.e., phonemes) (Ladefoged & Johnson, 2010). For instance, phonemic contrasts [d] vs. [t], [d] vs. [b], and [b] vs. [m] are distinguished by voice onset time (VOT, voiced vs. voiceless), place of articulation (tongue vs. lip), and manner of articulation (plosive vs. nasal), respectively. Physiologically, they are determined by articulatory gestures of the vocal folds in the larynx and articulators in the vocal tract (Browman & Goldstein, 1989). Ultimately, these gestures are controlled by the speech motor system in the brain, including the motor cortex with effector-specific somatotopy (i.e., dissociative lip, tongue, and larynx representation regions) (Brown et al., 2008). Intriguingly, speech perception has been postulated to recruit motor simulation by the speech motor system. The motor theory of speech perception proposes that categorical perception of phonemes (i.e., mapping varying acoustic features onto invariant phonological units) is achieved by reconstructing their articulatory motor sequences (Liberman et al., 1967) (see illustration in **Figure 1A**). Lesion studies (Basso et al., 1977; Blumstein et al., 1977) and functional neuroimaging studies (Watkins & Paus, 2004; Zatorre et al., 1992) have confirmed that the frontal cortex subserves such articulatory assistance function in speech perception. According to the dual-stream models of speech processing (Hickok et al., 2011; Hickok & Poeppel, 2007; Pulvermüller & Fadiga, 2010), the articulatory network belongs to the dorsal stream that maps phonological analyses onto motor representations, i.e., sensorimotor integration. Compared with the ventral stream-based classical models of speech comprehension (Tremblay & Dick, 2016), the dual-stream models emphasize a reciprocal auditory-motor transformation mechanism that supports the centrality of action in perception (i.e., embodiment, Wilson, 2002). Moreover, it accounts for human listeners' outstanding perceptual ability under adverse listening conditions (Du et al., 2014, 2016; Du & Zatorre, 2017), and has provoked a wide range of research interest in clinical studies (Behroozmand et al.,

2018) and natural language processing research (Anumanchipalli et al., 2019).

In recent years, TMS studies have revealed a causal role of motor cortex in categorical speech perception in a manner of fine-grained somatotopy where a lip motor subregion engages in categorizing bilabial consonants (D'Ausilio et al., 2009; Möttönen et al., 2013; Möttönen & Watkins, 2009), whereas a tongue motor area is recruited in distinguishing dental consonants (D'Ausilio et al., 2009; D'Ausilio, Bufalari, Salmas, & Fadiga, 2011; Fadiga et al., 2002) and vowels (D'Ausilio et al., 2014). However, compared with the tongue and lip motor areas, the contributions of the LMC have rarely been investigated. Studies in articulatory phonetics and neurophysiology have shown that the LMC regulates vocal pitch and controls voicing (Simonyan, 2014) by inducing changes in the tension and adduction/abduction of the vocal folds (Gick et al., 2013). This allows tonal language (e.g., Mandarin) speakers to produce lexical tone and consonant VOT contrasts, respectively. Following the logic of the somatotopic motor involvement in speech perception, we hypothesized that the LMC is causally involved in categorizing phonemic contrasts characterized by pitch contour and VOT, corresponding to its indispensability in producing these features. Take one step further, if the LMC is involved, it is unclear how the left and the right LMC are differently weighted in categorical speech perception, and how the involvement is modulated by task difficulty. In terms of lateralization, although it is widely accepted that phonological processing is left-lateralized, especially for the dorsal pathway (Hickok & Poeppel, 2007; Liang & Du, 2018), studies have also revealed the engagement of the right dorsal pathway in processing human voice (Lévêque et al., 2013) and prosody (Sammler et al., 2015). Although the LMC is hypothesized to encode both phonological (e.g., VOT) and prosodic (e.g., pitch) cues, the lateralization of its involvement remains elusive. In terms of the effect of task difficulty, debates still remain regarding whether the motor engagement is essential in identification of speech features (Wilson, 2009), or it is complementary and only necessary when cognitive demands increase (Hickok et al., 2009). Answering the above questions will promise a deeper and more comprehensive understanding of the sensorimotor circuits in speech processing.

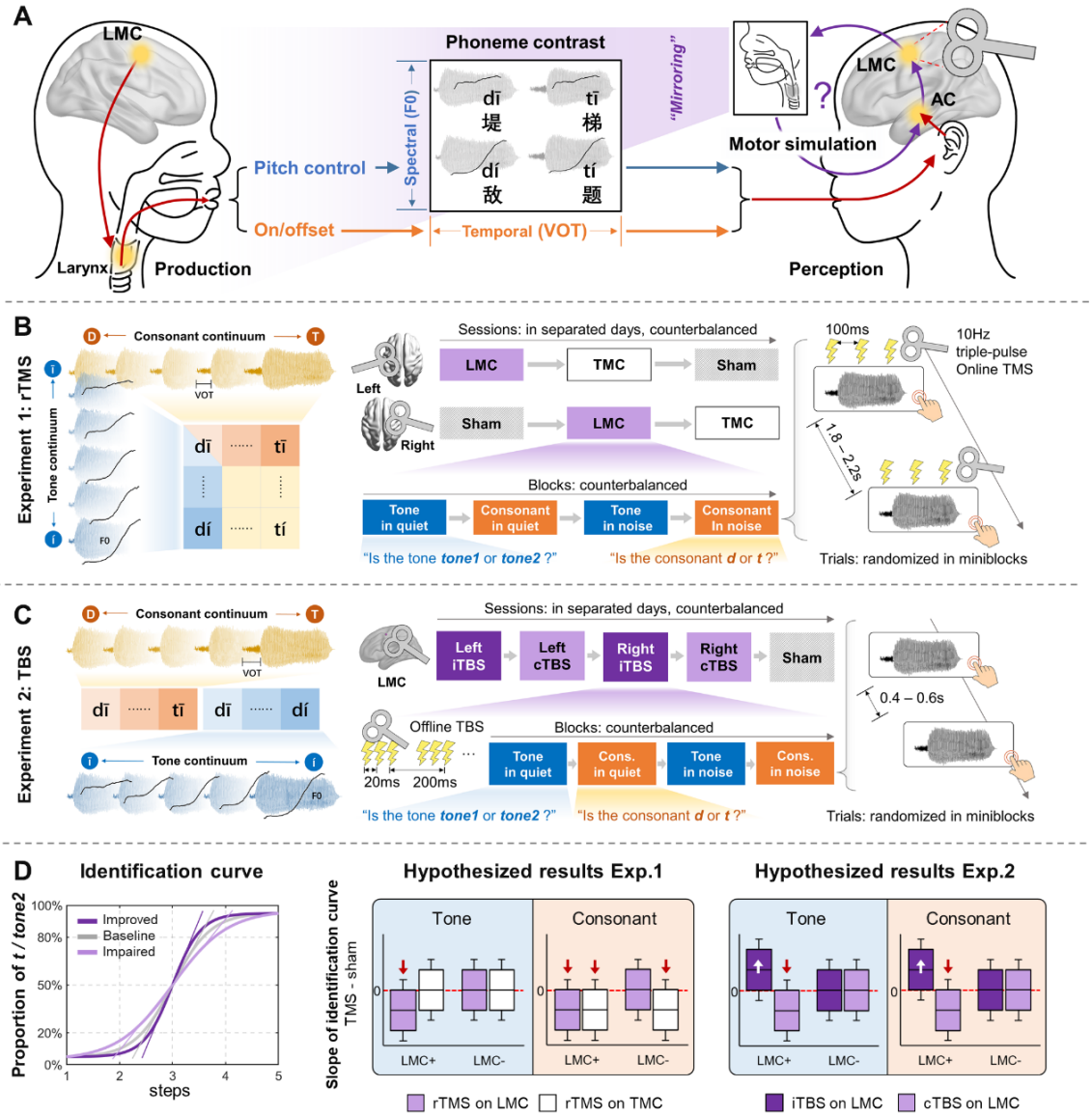


Figure 1. Experimental design.

(**A**) Schematic illustration of sensorimotor integration in Mandarin speech perception and protocol of this study. The LMC commands the larynx to control the pitch of voice and the onset and offset of phonation during speech articulation. Correspondingly, the LMC is assumed to engage in the categorical perception of lexical tone and consonant (VOT contrast) in Mandarin listeners, in a way of motor simulation that "mirrors" its function in speech production.

(**B**) Design of Experiment 1 (rTMS experiment). Left shows the orthogonal design of the tone–consonant stimuli matrix. Middle and right show the experimental procedure at session, block, and trial levels. Note that, we used Pinyin (Huang & Liao, 2017), the official transcription system for Mandarin Chinese, in task instructions (middle). Here, tone1 and tone2 refer to the high-level tone (55) and the mid-rising tone (35), respectively; "d" and "t" refer to the dental plosive [t]

with shorter VOT (unaspirated) and the dental plosive [t^h] with longer VOT (aspirated), respectively. 10 Hz triple-pulse online rTMS was applied at the syllable onset in each trial.

(**C**) Design of Experiment 2 (TBS experiment). Left shows the separate lexical tone and consonant continua. Middle and right show the experimental procedure at session, block, and trial levels. Instructions were identical to Experiment 1. Offline intermittent TBS (iTBS) or continuous TBS (cTBS) was applied to the left or the right LMC before the task.

(**D**) Hypothetical results. Left shows how TMS may affect curves of identification: compared with baseline (grey line), inhibitory stimulation (rTMS and cTBS) would "flatten" the curve and impair performance (light purple line), while excitatory stimulation (iTBS) would "steepen" the curve and improve performance (dark purple line). Oblique lines crossing the point of subjective equality (PSE) represent slopes of the curves. Middle and right show how slopes of tone and consonant identification curves may be altered by TMS in Experiment 1 (middle) and Experiment 2 (right) if (+) or if not (-) the LMC is engaged in Mandarin categorical perception, compared with sham baseline (red dashed lines). LMC+/-: when LMC is or is not hypothesized to be involved. Red arrows: inhibitory effect; white arrows: excitatory effect. The boxplots in the middle and right portions are schematic representations of possible slope distributions.

For the whole figure: blue and orange represent tone contrast (panel A)/identification tasks (panel B, C, and D) and consonant contrast (panel A)/ identification tasks (panel B, C, and D), respectively; light purple, dark purple, white, and grey represent inhibitory TMS (rTMS and cTBS), excitatory TMS (iTBS), control site (TMC stimulation), and sham (panel B and C) and their presumable effects (panel D), respectively. These color patterns are the same in **Figure 2** and **Figure 3**.

LMC: Laryngeal motor cortex; TMC: Tongue motor cortex.

Here, we delivered repetitive TMS (rTMS, Experiment 1) or theta-burst stimulation (TBS, Experiment 2) to participants' left or right LMC to investigate if categorization performance of Mandarin lexical tone and that of consonant with VOT contrast would be modulated (**Figure 1D**, middle and right). We used Mandarin Chinese because it is a typical tonal language where pitch contours are treated as phonemes to determine meaning (i.e., tonemes) (Chao, 1968). We used TMS for its relatively high spatial resolution (Bolgnini & Ro, 2010) and the ability to estimate the redundancy of cortical regions in cognitive processes (Ziemann, 2010). And importantly, we systematically compared the left and right LMC stimulation across

subjects (Experiment 1) or within subjects (Experiment 2) to probe the hemispheric asymmetry of the LMC engagement. In addition, we added masking noise in half of the trials to investigate if the involvement of the LMC would be upregulated by task demands and its interaction with lateralization.

Results

Experiment 1

In Experiment 1, we used rTMS to estimate changes in Mandarin tone and consonant categorical perception in a group of 64 young adults (Materials and methods, Participants). Participants were separated into two matched groups for left and right hemisphere stimulation, respectively (Materials and methods, Experiment 1, TMS paradigm). Experiment 1 included 3 rTMS sessions: one sham session, one LMC stimulation session, and one session for tongue motor cortex (TMC) stimulation. The TMC stimulation was set as a site control in verifying the paradigm as its effect has been well recognized (D'Ausilio et al., 2009; D'Ausilio, Bufalari, Salmas, & Fadiga, 2011; Fadiga et al., 2002). The MNI coordinates of the LMC [±40, -5, 50] and the TMC [±59, -3, 36] were defined by a functional localization task in 48 Mandarin speakers (see Appendix, Functional localization experiment). Within each session, participants underwent four blocks of syllable identification tasks: tone and consonant tasks with or without noise masking. The order of sessions and blocks was counterbalanced across participants in a Latin square procedure. In each trial, participants made speeded tone (tone1 vs. tone2) or consonant (d vs. t) two-alternative forced-choice judgements while ignoring the other dimension (consonant/tone) after listening to a syllable. Began with the syllable onsets, 3 pulses of 10 Hz biphasic rTMS were applied to the LMC or the TMC (Materials and methods, Experiment 1, Procedure, Stimuli presentation and **Figure 1B**, middle and right panels for details of experiment procedure). Different from offline high-frequency rTMS, 10 Hz online rTMS protocol has been frequently shown to inhibit cognitive processing (Beynel et

al., 2019) and especially down-regulate speech perception (Hartwigsen et al., 2016).

Syllables were randomly selected from a 5×5-step tone–consonant continuum matrix (**Figure 1B**, left panel). The matrix was generated from four Mandarin syllables ([ti55], [ti35], [t^hi55], and [t^hi35], corresponding to "堤" (riverbank), "敌" (enemy), "梯" (ladder), and "题" (title), respectively, by orthogonally synthesizing VOT (from the unaspirated [t] to the aspirated [t^h]) and F0 continua (from the high-level tone to the mid-rising tone). To create a VOT continuum, short segments of aspiration noise were stepwise appended between the release-burst and the phonation onset of the unaspirated [t] until reaching the aspirated [t^h]. To make an F0 continuum, a pair of syllables with tone contrast and shared identical VOT were entered into STRAIGHT synthesis algorithms (see Materials and methods, Experiment 1, Stimuli for details). To increase sensibility to rTMS, matrices were individualized to optimally cover each participant's range of perceptual ambiguity (see Materials and methods, Experiment 1, Stimuli presentation for details).

Slopes of psychometric functions served as the response variable. The lower the slopes, the weaker the participant's capacity to categorize phonemes (see Fitting psychometric function, Preprocessing of the slope, and Stimulation effect analyses for Experiment 1 in Materials and methods and **Figure 1D** left panel). Unexpectedly, we found competitions between tone and consonant perception such that slopes were affected by the ambiguity of the unattended dimension (see Appendix, Competitions between consonant and tone perception). To rule out this interference, for each block, slopes were separately extracted for trials with syllables being unambiguous, half-ambiguous, and ambiguous in the orthogonal dimension.

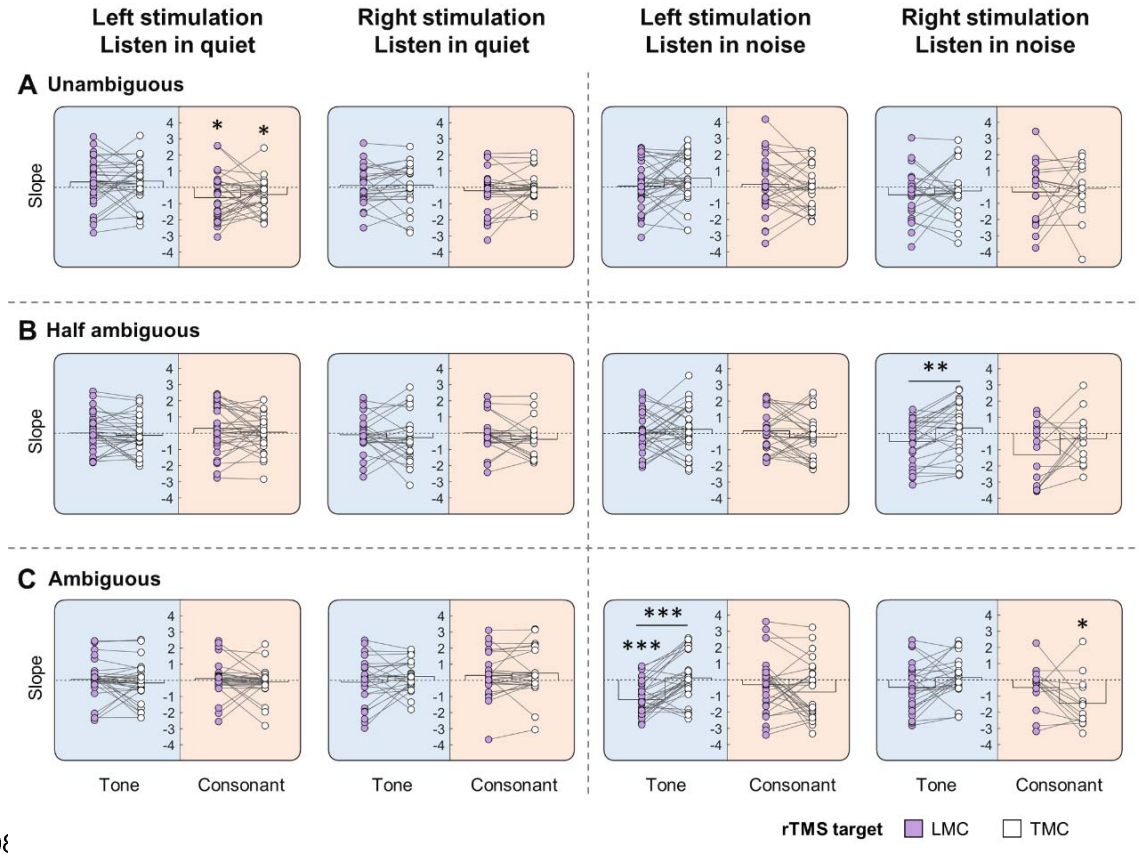


Figure 2. rTMS effects on tone and consonant categorical perception in Experiment 1.

(**A**) When syllables were unambiguous in the unattended dimension, rTMS upon the left LMC and TMC impaired consonant perception in quiet.

(**B**) When syllables were half-ambiguous in the unattended dimension, compared with right TMC stimulation, rTMS upon the right LMC impaired tone perception in noise.

(**C**) When syllables were ambiguous in the unattended dimension, rTMS upon the left LMC impaired tone perception in noise while stimulating the right TMC impaired consonant perception in noise.

For each panel, dots represent log slopes of psychometric curves for each participant; bars show the group average of slopes in each task. Light purple dots represent slopes in conditions where LMC was stimulated, whereas white dots are for slopes in TMC stimulation conditions. The blue background represents tone tasks, whereas the orange background represents consonant tasks.

LMC: Laryngeal motor cortex; TMC: Tongue motor cortex.

Statistical tests were performed by non-parametric tests comparing rTMS effects with zero, and comparing rTMS effects upon LMC with TMC within the same tasks. P values were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

Results showed that rTMS upon the LMC exerted inhibitory effects on both tone and consonant categorical perception. Tone perception in noise masking was impaired by stimulating the left LMC when the unattended consonant dimension was ambiguous (compared with sham: $p_{fdr} < 0.001$, Cohen's d = -1.166, permutation test; compared with TMC stimulation, Z = -4.027, $p_{fdr} < 0.001$, Cohen's d = -1.086, Wilcoxon signed-rank test, **Figure 2C**), and by stimulating the right LMC when the unattended consonants were half-ambiguous (compared with TMC stimulation: Z = -3.175, $p_{fdr} = 0.009$, Cohen's d = -0.838, Wilcoxon signed-rank test, **Figure 2B**). Left LMC stimulation also impaired consonant perception without noise-masking when tones were unambiguous (compared with sham: LMC, $p_{fdr} = 0.033$, Cohen's d = -0.459, permutation test, **Figure 2A**). Meanwhile, consistent with previous studies (D'Ausilio et al., 2009; D'Ausilio, Bufalari, Salmas, & Fadiga, 2011; Fadiga et al., 2002), perception of dental consonant without noise masking (in quiet) was impaired by rTMS upon the left TMC when tones were unambiguous (compared with sham: $p_{fdr} = 0.033$, Cohen's d = -0.467, permutation test, **Figure 2A**), and by rTMS upon the right TMC when tones were ambiguous (compared with sham: $p_{fdr} = 0.040$, Cohen's d = -0.940, permutation test), confirming the effectiveness of our paradigm. In contrast, rTMS upon the TMC did not influence tone perception in any condition (ps > 0.05). This validated the focality of rTMS and that tone perception is effector-specific to the LMC.

Experiment 2

To replicate findings of LMC stimulation in Experiment 1, we used offline TBS in another group of 26 participants. This is a TMS paradigm whose directionality has been extensively explored as iTBS and cTBS increase and decrease cortical excitability, respectively (Suppa et al., 2016). Moreover, to avoid competitions between tone and consonant, here we used one F0 and one VOT continuum, instead of a matrix, for the tone and consonant task, respectively (Materials and methods, Experiment 2, Stimuli, **Figure 1C**, left panel). Furthermore, to directly estimate

hemispheric asymmetry of the LMC in categorical speech perception, each participant received TBS upon both hemispheres in five separated sessions: sham, iTBS on the left LMC, cTBS on the left LMC, iTBS on the right LMC, and cTBS on the right LMC (**Figure 1C**, middle panel). In each session, they performed four tasks resembling Experiment 1 after receiving TBS (**Figure 1C** middle and right panels; see Materials and methods, Experiment 2 for details of Experiment 2). Slopes of curves were analyzed as in Experiment 1. We hypothesized that iTBS and cTBS upon the LMC would enhance and impair perception, respectively (**Figure 1D**, right panel).

Results showed that, cTBS upon the LMC inhibited both tone and consonant categorical perception. Tone perception with noise masking was impaired by cTBS upon the left LMC (compared with sham: p_{fdr} = 0.029, Cohen's d = -0.688, permutation test, **Figure 3C**). Consonant perception was impaired by cTBS upon the left LMC both when noise masking was absent (compared with sham: p_{fdr} = 0.049, Cohen's d = -0.474, permutation test, **Figure 3A**) and present (compared with sham: p_{fdr} < 0.001, Cohen's d = -1.918, permutation test; compared with iTBS, Z = -3.715, p_{fdr} < 0.001, Cohen's d = -1.030, Wilcoxon signed rank test, **Figure 3C**). cTBS upon the right LMC also impaired consonant perception in noise (compared with sham: p_{fdr} < 0.001, Cohen's d = -2.323; in comparison with iTBS, Z = -3.901, p_{fdr} < 0.001, Cohen's d = -1.970, Wilcoxon signed rank test, **Figure 3D**) but did not affect consonant perception in quiet or tone perception (ps > 0.05). However, iTBS did not exert any effect in any condition (ps > 0.05).

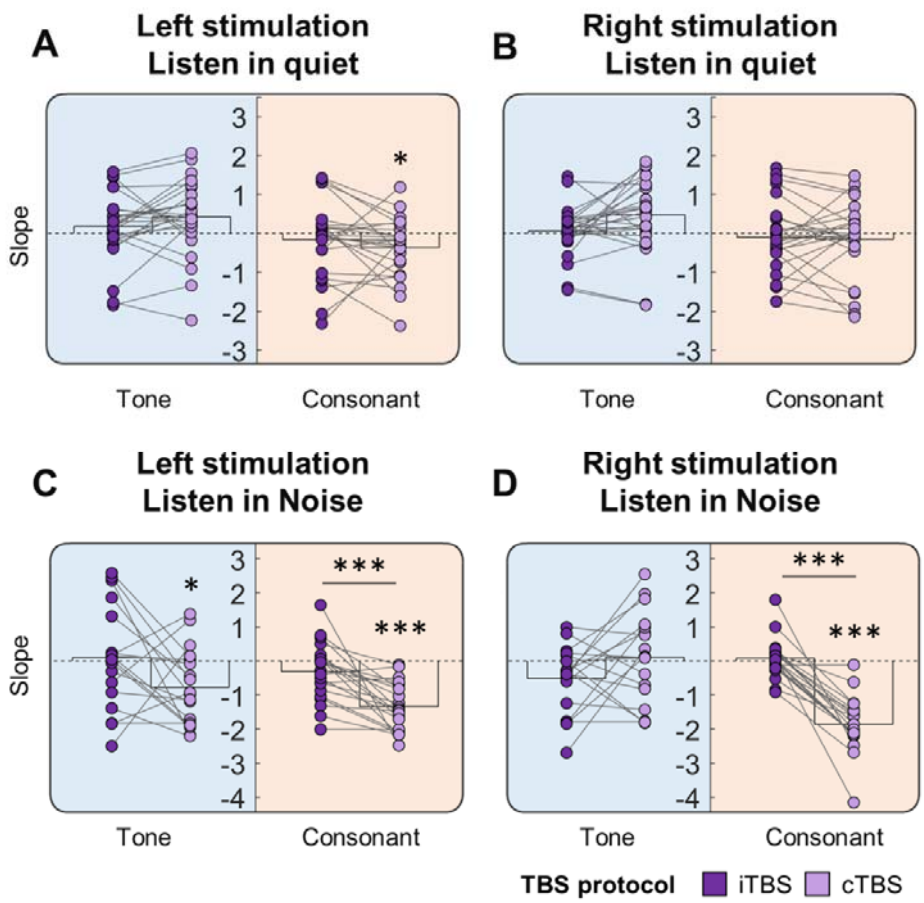


Figure 3. TBS effects upon the LMC on tone and consonant categorical perception in Experiment 2.

(**A**) cTBS upon the left LMC impaired consonant perception in quiet.

(**B**) TBS upon the right LMC had no effect on tone or consonant perception in quiet.

(**C**) cTBS upon the left LMC impaired both tone and consonant perception in noise.

(**D**) cTBS upon the right LMC impaired consonant perception in noise.

See Figure 2 legend for a detailed description of color patterns.

Statistical tests were performed by non-parametric tests comparing TMS effects with zero, and comparing effects of cTBS with iTBS within the same tasks. P values were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

Discussion

Human LMC for lexical tone and consonant categorization

Corresponding to the motor theory of speech perception (Liberman et al., 1967), a

dorsal pathway for speech production has been postulated to facilitate perception (Hickok et al., 2011; Hickok & Poeppel, 2007; Pulvermüller & Fadiga, 2010). Lesion studies gave early causal evidence as patients reported impaired perception of voicing in consonant (Blumstein et al., 1977), lexical tone (Gandour et al., 1992), and prosody (Witteman et al., 2011) following damage to the frontal areas. More recently, "virtual lesion" studies by TMS have shown that sensorimotor integration in speech lies in the effector-specific somatotopy (D'Ausilio et al., 2009; Schomers et al., 2015). However, whether and how the larynx motor area is recruited in categorical speech perception remains elusive. Here we provide causal evidence for shared representations of the LMC in classifying upcoming speech with its roles in pitch regulation and voicing control during articulation.

TMS studies have reported that the LMC is causally involved in discrimination of vocal pitch (D'Ausilio, Bufalari, Salmas, Busan, et al., 2011), singing voice (Lévêque et al., 2013), and paralinguistic prosody (Sammler et al., 2015). A recent study demonstrated that lexical tone processing was suppressed after stimulating the left motor cortex (Tang et al., 2021). These pioneering works pointed to a causal role of the LMC in categorizing linguistic pitch contours (i.e., lexical tones), which is for the first time confirmed by the present study. In addition, we provide substantial evidence that the LMC is causally recruited in detecting subtle VOT differences for consonants. This parallels its motor function to accurately control the moment of voicing onset, an acoustic feature crucial for voicing contrasts (Ladefoged & Johnson, 2010).

Nevertheless, the actual location of the LMC is still debated. Although in this study we targeted on a hypothetical dorsal LMC (Brown et al., 2008), the human homolog of the non-human primate LMC may be located in the ventral motor cortex (Belyk & Brown, 2017). From the evolutionary perspective, a human-specific dorsal LMC has emerged to handle vocal learning and gesture control subserving language communications (Jarvis, 2019), through a hypothetical "duplication and migration" process (Mars et al., 2018). A similar dorsal vocal motor system was also found in songbirds to support the acquisition of complicated vocal sequences (Jarvis, 2019), sharing a convergent genetic basis with human (Pfenning et al., 2014). In sum, a

dorsal human LMC may implement fine-grained computations of speech vocalization (Eichert et al., 2020). However, recent studies have also found that the ventral LMC is engaged in vocal pitch production (Dichter et al., 2018), and voicing processing (Tamura et al., 2022). Future studies are expected to directly compare causal contributions of the dorsal and ventral LMC to categorical speech perception. In addition, the position of the dorsal LMC varies across subjects and studies, and is in close proximity of the lip representation (Brown et al., 2008; Eichert et al., 2020). We used dental consonants here to avoid confounds related to lip motor processing. Also, we used activated coordinates from a functional localization task performed by Mandarin speakers to rule out potential anatomical differences shaped by language experience (see Appendix, Functional localization experiment). Despite these efforts, future studies taking reliable individual motor maps as stimulation targets are still expected.

Left-lateralized involvement of the LMC

Table 1. TMS effects on tone and consonant categorical perception pooling Experiments 1 and 2.

Stimulation pattern	Difficulty		Contrast
	Noise masking	Ambiguity	
Stimulating the left LMC, tone tasks			
TBS	+	Unambiguous	cTBS-Sham
rTMS	+	Ambiguous	Larynx-Sham
rTMS	+	Ambiguous	(Larynx-Sham)-(Tongue-Sham)
Simulating the left LMC, consonant tasks			
rTMS	-	Unambiguous	Larynx-Sham
TBS	-	Unambiguous	cTBS-Sham
TBS	+	Unambiguous	(cTBS-Sham)-(iTBS-Sham)
TBS	+	Unambiguous	cTBS-Sham
Stimulating the right LMC, tone tasks			
rTMS	+	Half ambiguous	(Larynx-Sham)-(Tongue-Sham)
Stimulating the right LMC, consonant tasks			
TBS	+	Unambiguous	(cTBS-Sham)-(iTBS-Sham)
TBS	+	Unambiguous	cTBS-Sham

As summarized in **Table 1**, the involvement of the LMC in speech categorical

perception is biased to the left hemisphere. For lexical tone perception, stimulating the left LMC tended to more effectively block task performance (rTMS ambiguous conditions and cTBS), compared with right LMC stimulation (rTMS half-ambiguous conditions). For consonant perception, performance in tasks without noise-masking was only hampered by left but not right LMC stimulation (**Figure 2**; **Figure 3**; **Table 1**). Note that, hemispheric asymmetries were also modulated by the task difficulty, but this interaction is discussed in the next section. Here we firstly address the broad issue of hemispheric asymmetries.

Lexical tone is a suprasegmental linguistic cue sharing identical articulatory acoustic essence with prosody and congruent language functions with segmental phonemes (Liang & Du, 2018). The uniqueness of lexical tone has motivated research and fierce debate on its hemispheric asymmetry. Evidence has been provided for left-dominance (Liang & Du, 2018; Tang et al., 2021), right-dominance (Ge et al., 2015), and bilateral activations (Chien et al., 2020; Si et al., 2017) in tone processing. The current study brings new causal evidence for left superiority, specific to the LMC. Notably, Tang et al.(2021) discovered that tone perception in tonal and nontonal language speakers was separately hampered by the left and right motor cortex stimulation. Hence, it is possible that tonal language experience drives the motoric hub for tone perception from right to left (Zatorre & Gandour, 2008). On the other hand, prosodic pitch analyses recruit bilateral dorsal pathways for articulatory simulation with left-specialized phonological representations (Chien et al., 2020). Future cross-language studies investigating how tonal language experience (tonal vs. nontonal language speakers) and linguistic function of pitch contours (linguistic lexical tone vs. paralinguistic prosody) modulate hemispheric asymmetry of the LMC would be informative regarding the universality of our discovery. Also, further neuroimaging studies with connectivity analysis are needed to delineate the cerebral network and its processing asymmetry where bilateral LMCs interact with auditory cortices.

In terms of consonant, consistent with the prevailing left-dominance models for segmental phoneme perception (Poeppel, 2003; Zatorre et al., 2002), we found that

motor simulation of voicing in consonants is biased to the left LMC. Indeed, a left-lateralized speech motor cortex eases intra-hemispheric auditory sampling of phonemes (Möttönen et al., 2013). However, stimulating the right LMC also inhibited consonant perception when noise was presented (Experiment 2). To our knowledge, this is the first study to report the participation of the right dorsal stream in categorizing fast-changing segmental linguistic units.

Most studies of brain asymmetry in speech perception have been concentrated on functional distributions of the ventral pathways as the left auditory cortex is better tuned for segments with higher temporal modulation rates (i.e., linguistic cues such as phonemes), while the right is better at capturing cues with higher spectral modulation rates (i.e., prosody) (Albouy et al., 2020; Flinker et al., 2019). Hemispheric structural and functional differences of auditory cortices may account for this asymmetry (the *intrinsic* factors) (Poeppel, 2003; Toga & Thompson, 2003; Zatorre et al., 2002). In addition, individual differences in auditory asymmetry have been ascribed to the strength of auditory-frontal connectivity (Poeppel & Assaneo, 2020) (the *external* factors). Indeed, resembling the ventral pathways, latest studies have also demonstrated asymmetric dorsal pathways where sensorimotor integration in segmental linguistic units is left-dominant (Hickok & Poeppel, 2007; Liang & Du, 2018; Tang et al., 2021) whereas that for the suprasegmental paralinguistic prosody is biased to the right hemisphere (Sammler et al., 2015; Tang et al., 2021). Thus, it is speculated that the auditory asymmetry in speech perception may also be driven by the motor heterogeneity (Tang et al., 2021).

Nonetheless, driving auditory asymmetry does not naturally mean that the dorsal pathways share the same hemispheric specializations, as the corpus callosum subserves inter-hemispheric interactions (Bloom & Hynd, 2005). Studies have reported that neural oscillations in bilateral frontoparietal cortices couple with the left auditory cortex in speech perception (Park et al., 2015). Besides, particularly for the motor cortex, the right LMC is essential in detecting human voice (Lévêque et al., 2013), and the right motor cortex entrains speech (Hincapié Casas et al., 2021) and is involved in auditory-speech coupling (Keitel et al., 2017) resembling its left

counterpart. Similarly, in the current study, we showed that despite the general left-lateralization, the right LMC was essential in both lexical tone and consonant perception in noisy conditions. Therefore, it is evident that the right motor cortex is also engaged in sensorimotor integration in speech perception.

Nevertheless, since speech perception incorporates several steps (McClelland & Elman, 1986), current findings do not neglect influences on auditory asymmetry exerted by a knowledge-based left-biased frontoparietal language system (Pylkkänen, 2019). Instead, it calls for combinatory works of hierarchical language mapping (Meyer, 2018) and multi-stage neural network approaches (Güntürkün et al., 2020) to illustrate how functions and asymmetries of bilateral dorsal pathways temporally expand, and how they interplay with ventral pathways.

Modulatory role of the LMC

Whether sensorimotor integration is essential (Wilson, 2009) or modulatory (Hickok et al., 2009) in speech perception remains controversial. The current study supports a modulatory view as increased task difficulty (i.e., adding noise) prompted more extensive motor engagement, both for tone and consonant tasks (see **Table 1**). This coincides with previous neuroimaging (Du et al., 2014, 2016) and TMS (Nuttall et al., 2017) studies. Notably, the consonant results are inconsistent with Sammler et al.(2015) as they found that stimulating the LMC in neither side affected consonant perception. However, our tasks were more challenging due to extra noise masking, and Mandarin VOT contrasts may naturally be more confusable than English (no significant voiced plosives are in Mandarin (Huang & Liao, 2017)). Our findings demonstrate that sensorimotor integration functions as a complementary mechanism to counteract degraded listening (Hickok et al., 2011). However, as we adopted acoustic continua that were difficult to classify per se, here we could not estimate motor engagement in categorization of natural, clear speech.

In addition, we found possible interactions between task difficulty and hemispheric asymmetry. For tone tasks, as was discussed above, inhibitory effects

were mostly triggered by left LMC stimulation, whereas stimulating the right LMC only hampered performance when masking noise and competitions from consonant (i.e., half-ambiguous) both existed. For consonant tasks, stimulating the left LMC affected both perception in quiet and in noise, but stimulating the right LMC only blocked perception in noise. In sum, the left LMC seems to be a key node for both tone and consonant categorical perception, whereas the right LMC is less robustly involved or is redundant (Friston & Price, 2003). When perception becomes extremely cognitive demanding, bilateral LMCs, especially the right LMC, are upregulated as a means of compensation. Nonetheless, how the left dorsal stream cooperates with its right counterpart to overcome perceptual challenges in speech perception requires further studies.

Limitations and conclusions

Despite the progress we made, limitations exist for the current study. For example, it is unclear to what extent the laboratory findings on syllable-level perception can be generalized to natural language communications. In other words, perception of continuous sentences with rhythmical, hierarchical, and contextual information, may require more sophisticated functions of the motor cortex such as temporal (Rimmele et al., 2018), motor (Pickering & Gambi, 2018), and contextual prediction (Brodbeck et al., 2022), as well as envelope tracking (Ding & Simon, 2012) in addition to categorical perception. Further studies are needed to probe the functional relevance of the LMC for speech perception in natural settings.

Furthermore, our work is limited in terms of the neurobiological infrastructure for both causal inference for neurostimulation and dynamic neural network delineation. For the former, after stimulation, short-term functional reorganization of neural networks may occur to counteract downregulation of the target (Bergmann & Hartwigsen, 2020) and complicate the causal interpretation. For the latter, the stimulation protocol we used provided limited temporal information to dissociate stages of motor involvement. Also, we were not able to probe the functional interplay

among bilateral dorsal LMCs and the ventral pathways during speech perception (e.g., Tamura et al.(2022)). Future studies combining brain stimulation and neuroimaging techniques are needed to fill these gaps.

In conclusion, this study is the first to demonstrate that the LMC is causally involved in categorical perception of lexical tone and consonant, which is grounded in its articulatory functions (i.e., pitch and voicing control). Moreover, its engagement seems to shift from left-dominant to bilaterally symmetric as task demands increase. This study provides new evidence to support the motor theory of speech perception, and more generally, active perception and embodied cognition.

Materials and methods

Participants

64 young participants (36 females, mean age = 22.19 years, SD = 2.88) took part in Experiment 1. Participants were divided into two matched groups for left and right hemisphere stimulation, respectively (Materials and methods, Experiment 1, TMS paradigm). For each group, the group size (left: 35; right: 29) is sufficient to detect medium to large effects as estimated by G*Power3. Another 26 participants (14 females, Mean age = 22.38 years, SD = 2.88) took part in Experiment 2. The group size is sufficient to detect medium to large effects as estimated by G*Power3. All participants were right-handed as assessed by Edinburg Handedness Inventory (mean laterality quotient: Experiment 1: 80.34, SD = 22.36; Experiment 2: 80.38, SD = 18.91) (Oldfield, 1971), and had normal hearing as measured by the Bell Plus audiometer (Bell, Inventis, Italy) with a TDH-39 headphone at 0.125, 0.25, 0.5, 1, 2, 4, and 8 kHz (≤ 20dB SPL). None of them had self or family history of neurological, traumatic, or psychiatric diseases. Participants were Mandarin Chinese speakers, and were all non-musicians (no more than 2 years of music training) as assessed by the Chinese version of Montreal Music History Questionnaire (MMHQ) (Coffey et al.,

2011). All participants gave written informed consent prior to the experiment. They were paid ￥50 per hour for participation in behavioral sessions, and ￥100 per hour for TMS sessions. The experiment was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences.

Methods details

Experiment 1

Stimuli

Syllables were recorded from a young female Mandarin speaker (age = 24). Sounds were recorded in a sound-proof studio, by RODE NT1 microphone connected to an Audient ID14 soundcard (sampling rate = 44.1 kHz). Syllables were [ti55], [ti35], [t^hi55], and [t^hi35] (transcribed by the International Phonetic Alphabets (Ladefoged & Johnson, 2010) and Chao's system of "tone-letters" (Chao, 1980)), corresponding to Mandarin syllables "堤 dī" (riverbank), "敌 dí" (enemy), "梯 tī" (ladder), and "题 tí" (title) (marked by Chinese Phonetic Alphabets "Pinyin" (Huang & Liao, 2017)), respectively (**Audio S1**). Four syllables were in good sound quality and similar in volume and shape of envelope. Pitch contours of each pair of syllables with the same tone were similar and normalized according to Chao's system. Environmental noise was filtered by Praat 6 algorithms.

Each syllable was around 400 ms with an identical length of phonation portion. VOT of the aspirated phoneme [t^h] was about 80 ms. Four syllables were 5 kHz low-passed filtered by Praat 6 algorithms and matched according to the average root mean square sound pressure level (SPL). To control acoustic features, four syllables were cut and re-concatenated in order to share similar phonation portions with varying preceding aspiration noise and overlaid pitch contours (see **Appendix Table 1** for acoustic properties of the syllables).

Continua were synthesized by Matlab R2016a and legacy-STRAIGHT. Each

continuum included multiple steps of syllable, which varied in only one acoustic property (VOT or F0 contour). Two types of continua were used: consonant continua ranging from [t] to [t^h] ("d" to "t" in Pinyin; note that the "d" in Pinyin is the unaspirated voiceless dental plosive, instead of the voiced [d] (Huang & Liao, 2017)), as well as tone continua ranging from high-level tone [55] to mid-rising tone [35] (tone1 to tone2). To generate a VOT continuum, implemented by manual code in Matlab, short segments (1/54 the length of natural VOT) of aspiration noise were stepwise appended between release-burst and the phonation onset of the unaspirated [t] until reaching [t^h]. Also, to keep the authenticity of syllables, phonation periods of a VOT continuum were also morphed from [ti] to [t^hi] using STRAIGHT synthesis algorithms, but this could hardly affect categorical perception since VOT is the most significant feature to distinguish these two consonants. To make an F0 continuum, a pair of syllables differing only in tone (i.e., one with high-level tone and another one mid-rising tone, but sharing the same VOT) were entered into STRAIGHT synthesis algorithms. Further, a 54×54-step matrix of tone–consonant continua was generated by orthogonal morphing. Hence, each syllable in the matrix represented a step in both tone and consonant continua.

Individualized matrix of stimuli

5×5-step individualized continuum matrices were generated to fit participants' ranges of perceptual ambiguity (steps in a continuum where participants were unable to clearly determine). To estimate the ranges for tone and consonant continua, in a preliminary experiment, participants performed tone or consonant identification tasks with a 9×9-step matrix selected from the 54×54-step matrix, using the method of constant stimuli. In each trial, participants heard a syllable, and needed to identify to which tone/consonant category the syllable belonged, ignoring the other dimension. Edges of ranges were manually detected. Based on the estimated edges, individualized continua were morphed as was described above (see **Appendix Table 1** for acoustic properties of the individualized continua, **Audio S2** for F0 continuum,

Audio S3 for VOT continuum).

Before the preliminary experiment started, participants were familiarized with the four natural syllables by listening to each of them. Then, they finished a 20-trial four-alternative forced-choice practice task of syllable identification. A correct rate of more than 90% (i.e., > 18 in 20 trials) was the requirement for entry. Afterwards, to ensure that participants understood the tasks, they finished a discrimination task to determine if the heard pair of syllables shared identical tone/consonant category. The discrimination task was continued until participants discriminated all pairs correctly.

Stimuli presentation

Stimuli were played by a Sennheiser in-ear headphone (IE 60) connected to a Steinberg UR 242 amplifier. Intensity levels of stimuli were calibrated by Larson-Davis sound level meter (Model 831, Depew, NY). Syllables were presented by Psychtoolbox-3 in Matlab 2016a at 78 dB sound pressure level (SPL), loud enough to resist online TMS noise.

In conditions with noise-masking, speech-spectrum noise (SSN) started as a block started and was kept playing until the block ended. SSN was generated by modulating white noise with the average spectrum of 5000 sentences spoken by 10 young female Mandarin speakers (Wu et al., 2005), and 4 kHz low-pass filtered. During the presentation, the intensity of masking noise was adjusted to produce a signal-to-noise ratio (SNR) of -4 dB, a medium level that was shown to induce maximal sensorimotor integration in syllable decision (Du et al., 2014, 2016).

Magnetic resonance imaging (MRI) acquisition and processing

For each participant, a T1-weighted anatomical image for target mapping was acquired by a 3-Tesla Siemens Magnetom Trio scanner with a 20-channel head coil, using the magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2200 ms, TE = 3.49 ms, field of view = 256 mm, flip angle = 8°,

spatial resolution = 1×1×1 mm). DICOM images were converted to NIfTI format by dcm2nii in MRIcron.

Procedure

Participants finished Experiment 1 on five separate days: day 1 for filling questionaries (informed consent, handedness inventory, and music history questionnaire), assessing hearing and TMS thresholds, and preliminary experiments; day 2 for MRI acquisition; day 3 to 5 for rTMS experiments. rTMS experiments included 3 sessions: one sham session, one LMC stimulation session, and one TMC stimulation session. The LMC and TMC targets were defined by the group-level MNI coordinates of geometric middle points of BOLD activations in a functional MRI localization experiment, where 48 participants performed larynx-related (say [a]) and tongue-related (say [t^h]) articulation tasks (LMC: [±40, -5, 50], TMC: [±59, -3, 36]; see Appendix, Functional localization experiment). Each of the three sessions was conducted separately by at least 3 days. The order of sessions was counterbalanced across participants using a balanced Latin square procedure.

Within each session, participants underwent four blocks of syllable identification tasks: tone/consonant tasks with/without noise masking. All four blocks used the same individualized 5×5-step matrix. The order of four blocks was counterbalanced across participants in a Latin square procedure. One block contained 2 mini-blocks. Each of the mini-blocks randomly presented all syllables in the 5×5-step tone–consonant matrix. To enhance the reliability of the fitted slope changes, in each mini-block, additional trials around the PSE were presented. This yielded a total of 78 trials with 39 trials in each mini-block. In each trial, participants heard a syllable while receiving online rTMS. After a syllable was presented, participants used their index and middle finger to press the “<” or “>” keyboard key to make a speeded judgement (to which tone/consonant category the heard syllable belonged), and ignore the other dimension. Participants always used the hand ipsilateral to the stimulated hemisphere to avoid

response bias induced by stimulating the hand motor cortex. The order of keys was balanced across participants. In each trial, a resting period of maximal 5 seconds after sound onset was left for a response. The next trial started 2s on average (1.8 – 2.2s, 10 ms step, uniformly distributed) after participants' response or the automatic end of a trial. No break time was left between adjacent mini-blocks.

To get prepared for the experiment, participants completed practice tasks before each session of the rTMS experiment. The practice task preceding the first rTMS session contained short tone/consonant identification tasks using a different continua matrix ([p]–[p^h] × tone1–tone2 in Mandarin), whereas the practice tasks before the second and third sessions included a short version of the formal experiment.

rTMS

Resting motor threshold estimation

Before the experiment, participants accepted 2 to 3 pulses of TMS at middle intensity levels to guarantee tolerance. Participants' resting hand motor threshold (RMT) was tested by placing the coil on the dorsal motor cortex localized by Brainsight 2 neuronavigation software and adjusting output intensity stepwise. The RMT was defined as the lowest stimulator output that induced at least 5 of 10 trials with visible hand motor responses. Participants' average RMT was 61.30% of maximal stimulator output (SD = 8.03%).

TMS paradigm

Participants were separated into two groups for left (N = 35, 20 females) and right (N = 29, 16 females) hemisphere stimulation, respectively. Two groups of participants were matched in age (left: mean = 22.43, SD = 3.03; right: mean = 21.90, SD = 2.67; $t(62) = 0.73$, $p = 0.47$), RMT (left: mean = 61.09%, SD = 7.69; right: mean = 61.55%, SD = 8.42; $t(62) = -0.23$, $p = 0.82$), and handedness (left: mean = 80.26, SD = 23.72;

right: mean = 80.44, SD = 20.60; t(62) = -0.03, p = 0.98).

During the experiment, triple-pulse rTMS was applied in synchronization with the syllables. Biphasic magnetic pulses were generated by a standard figure-of-eight coil from a Magstim Rapid2 stimulator (https://www.magstim.com/row-en/product/rapid-family/). The intensity of stimulation was set to 90% of participants' RMT. The frequency of stimulation was 10 Hz (100 ms between two pulses). The first pulse was always delivered at the syllable onset, followed by two subsequent pulses. Since a single TMS pulse induces aftereffects lasting for tens of milliseconds (Sandrini et al., 2011), a triple-pulse sequence of stimulation can shadow an entire syllable. Also, although it was suggested that high-frequency offline rTMS induces facilitatory effects (Fitzgerald et al., 2006), 10 Hz online rTMS protocol has been shown to inhibit cognitive processing (Beynel et al., 2019) and more specifically, down-regulate speech perception (Hartwigsen et al., 2016). Presumably, while offline TMS alters performance by inducing short-term neural plasticity (Thickbroom, 2007), online TMS affects cognition by modulating ongoing neural oscillations (Thut et al., 2011), implying that the frequency-dependent rule of offline rTMS effects may not fit online paradigms. To account for inter-subject variability in brain structure, MNI coordinates of the LMC and TMC targets were mapped to the participants' T1-weighted image with Brainsight2, and manually adjusted if coordinates were not in precentral gyrus while keeping the vertical axis positions constant. Positions of the coil were real-time tracked by the Brainsight system to keep the target being focused. The magnetic field was always kept perpendicular to precentral gyrus, heading towards the front of the participant. In sham conditions, the coil was placed on the vertex and the same stimulation sequences were delivered.

Experiment 2

Stimuli

Same syllables [ti55], [ti35], and [t^hi55], corresponding to Mandarin syllables "堤 dī"

(riverbank), “敌 dí” (enemy), and “梯 tī” (ladder) from Experiment 1 were used in Experiment 2. Different from Experiment 1, only one 54-step tone continuum ([ti55] to [ti35]) and one 54-step consonant continuum ([ti55] to [t^hi55]) were synthesized. Procedures of stimuli preprocessing and continuum morphing were identical to Experiment 1.

Individualized stimulus continua and SNR level

In Experiment 2, both ranges of perceptual ambiguity and SNR levels for noise masking were individualized. Participants first entered familiarization sessions identical to Experiment 1. To estimate individual ranges of perceptual ambiguity, a preliminary experiment similar to that in Experiment 1 was conducted, but participants performed tone and consonant identification tasks using tone and consonant continuum, respectively. Individualized SNR levels were estimated by the method of constant stimuli where participants performed tone and consonant identification tasks using unambiguous syllables ([ti55] and [ti35] for tone tasks, [ti55] and [t^hi55] for consonant tasks) masked by SSN. SSN (identical to Experiment 1 except 5 kHz low-pass filtered) began and ended simultaneously with the syllable, with additional 10 ms linear rise-decay envelopes. For each task, the SNR level where participants reached around 85% of decision accuracy was defined as the individualized SNR level: -16dB to 2dB in tone tasks; -12dB to 12dB in consonant tasks.

Stimuli presentation

Hardware and software to present stimuli in Experiment 2 were identical to that in Experiment 1. Syllables were presented at 65 dB SPL. This was lower than that in Experiment 1 since no TMS noise affected stimuli presentation.

In conditions with noise-masking, the intensity of SSN was adjusted to produce individualized SNRs.

MRI acquisition and processing

MRI acquisition and processing were identical to those in Experiment 1.

Procedure

Participants underwent five sessions: sham, iTBS on the left LMC, cTBS on the left LMC, iTBS on the right LMC, and cTBS on the right LMC. Each of the five sessions were conducted separately by at least 3 days. The order of sessions was counterbalanced across participants using a balanced Latin square procedure.

Within each session, participants first received one train of TBS or sham stimulation. After stimulation, participants underwent four blocks of syllable identification tasks similar to Experiment 1. Individualized tone and consonant continua were used for tone and consonant tasks, respectively. The order of four blocks was counterbalanced across participants in a Latin square procedure. One block contained 4 mini-blocks. Each of the mini-blocks randomly presented all syllables in the continuum. To enhance the reliability of the fitted slope changes, in each mini-block, additional trials around the PSE were presented. This yielded a total of 156 trials with 39 trials in each mini-block. Participants responded in the same way as in Experiment 1. In each trial, a resting period of maximal 5 seconds after sound onset was left for a response. The next trial started 0.5s on average (0.4 – 0.6s, 10 ms step, uniformly distributed) after participants' response or the automatic end of a trial. No break time was left between adjacent mini-blocks. Identical to Experiment 1, participants completed practice tasks before each TBS session.

TBS

Active motor threshold estimation

Participants' active motor threshold (AMT) was estimated with motor evoked potentials (MEPs) recorded from the right first dorsal interosseous muscle and were

automatically filtered by Brainsight 2. Here we used AMT instead of RMT because AMT has been confirmed to be a valid indicator of individualized TBS intensity levels (Mas-herrero et al., 2018). Note that, we found that, in Experiment 2, participants' RMT and AMT were highly correlated ($r = 0.857$, $p < 0.001$), showing that AMT captures similar individual differences to what RMT does. Before the experiment, participants accepted 2 to 3 pulses of TMS at middle intensity levels to guarantee tolerance. Participants' RMT was estimated as described in Experiment 1, except that the RMT was defined as the lowest stimulator output that induced at least 5 of 10 trials with MEP peak-to-peak amplitude greater than 50μV. Based on RMT, output intensity was adjusted to measure AMT. Participants were keeping a maximal voluntary hand contraction to a dynamometer during the measurement, and the AMT was defined as the lowest stimulator output that induced at least 5 of 10 trials with MEP peak-to-peak amplitude greater than 1000μV. This was similar to the criteria of 200μV when participants maintained 20% of maximal hand contraction (Huang et al., 2005), but was easier to control. 80% AMT was used as stimulation intensity during the experiment, but the intensity was always kept at least 40% of the maximal stimulator output.

TMS paradigm

During the experiment, participants received TBS after practice and before tasks. Biphasic magnetic pulses were generated by a standard figure-of-eight coil from a Magstim Rapid2 stimulator. The intensity of stimulation was set to 80% AMT or 40% of maximal stimulator output. We used the TBS protocol that delivered three 50 Hz pulses in each 200 ms period (Huang et al., 2005). For iTBS conditions, TBS was delivered for 2s followed by an 8s resting interval. A train of iTBS lasted for 192s and delivered 600 pulses in total. For cTBS conditions, 600 pulses of stimulation were continuously delivered for 40s. Navigation of the TMS coil was identical to that in Experiment 1. In sham conditions, the coil was placed over the left or the right LMC and delivered iTBS or cTBS, but was tilted so that only the edge was kept in touch

with the skin. In iTBS conditions, participants started to perform tasks immediately after the stimulation ended. In cTBS conditions, participants performed tasks 5min after stimulation to maximize TMS aftereffects (Huang et al., 2005).

Quantification and statistical analysis

Experiment 1

Fitting psychometric function

Scores of identification tasks were binarily coded: trials which participants judged as high-level tone (tone identification) or consonant [t] (consonant identification) were labeled as 0, and otherwise 1.

Logistic models were fitted using the non-linear *lsqcurvefit* function in Matlab R2016a, taking morphing steps in a continuum as the explanatory variable and scores as the response variable. The formula of the curve is shown below:

$$\hat{Y} = \frac{a - b}{1 + e^{-\sigma(X-\mu)}} + b \qquad \text{(Equation 1)}$$

$\hat{Y}$ is the estimated psychometric function (identification curve). X is the steps of a continuum. a and b are the upper and lower asymptotes of the curve. μ is the PSE. σ is the slope of the curve at the PSE.

To avoid overfitting, for each condition, upper and lower asymptotes (a and b) were calculated by an unconstrained fit on the average data, and were fixed for each block.

Preprocessing of the slope

For sham condition, slopes with negative values (< 0) or with PSE out of the continuum interval (< 1 or > 5) were eliminated as these parameters were mathematically invalid in the current context. For rTMS conditions, the "mathematically invalid" slopes could actually indicate that rTMS completely

disrupted categorical perception. Therefore, instead of being deleted, these invalid slopes were replaced by the floor values in the corresponding sham blocks. To get the floor values, outliers (3SD from mean) were removed from the preprocessed data for each condition. Then, the block-wise floor values were derived from the formula F = exp(mean(log(D)) - 2*SD(log(D))), where D was the data while F was the corresponding floor value. The preprocessed sham slopes and the floor value-adjusted rTMS slopes were logarithmically transformed to adjust for positive skewness (See Appendix figure 1 for a graphical illustration for this procedure).

Stimulation effect analyses

Since tone and consonant perception competed against each other as the stimulus ambiguity of the unattended dimension affected the slope of identification curve in the attended dimension (see Appendix, Competitions between consonant and tone perception), it is necessary to exclude such interference with rTMS effects. Therefore, for each block of tasks per session (LMC rTMS, TMC rTMS, and sham) per group of participants (left and right hemisphere stimulation), curve fitting, preprocessing, and estimation of rTMS effects were calculated separately for ambiguous (step 3), half-ambiguous (step 2 and 4), and unambiguous (step 1 and 5) trials according to the unattended attention (Appendix figure 2A).

rTMS effects were defined as differences between slopes in rTMS blocks and the corresponding sham blocks, and were statistically tested using single-tailed one-sample permutation test in Matlab 2016a (if rTMS - Sham < 0 or not). Compared with t-test, permutation test does not require normality assumption of sample distribution, as distributions of slopes in some conditions were skewed as estimated by Kolmogorov-Smirnov test of normality. To perform a permutation test, for each condition in rTMS blocks, we randomly selected participants and permuted their slopes in the rTMS block and sham block, contrasted rTMS to sham, and calculated the mean value. This permutation–contrast–average process was repeated 10^5 times and a permutation distribution of random mean differences was generated. P-value

was calculated by ranking the real rTMS effect among values in the permutation distribution: $p = (\text{ranking} + 1) / (10^5 + 1)$. In addition, the relative down-regulation of rTMS upon the LMC on categorical perception was estimated by comparing LMC stimulation with TMC stimulation effects on slopes ((LMC - Sham) - (TMC - Sham)) using single-tailed Wilcoxon signed-rank test in Matlab 2016a. Here, we used Wilcoxon signed-rank test instead of permutation test because the LMC stimulation and the TMC stimulation shared the same sham condition, and it was unable to simultaneously permute the sham condition with two rTMS conditions. To correct for multiple comparisons, for each group of trials (ambiguous, half-ambiguous, and unambiguous) in each group of participants (stimulating left or right hemisphere), FDR correction was used with p threshold = 0.05.

Experiment 2

Procedures of fitting psychometric function, preprocessing of the slope, and stimulation effect analysis were identical to those in Experiment 1. Nevertheless, trials were not grouped into different categories for curve fitting. In terms of TMS effect comparison, effects of cTBS (cTBS - Sham) and iTBS (iTBS - Sham) were compared by single-tailed Wilcoxon signed-rank test.

Acknowledgements

The authors would like to thank Xiaonan Li and Lei Zhang from the Institute of Psychology for help with neuroimaging data processing, Quan Yuan and Xinyuan Wan for data collection, and Xinyi Qian and Yi Weng for reviewing linguistics and phonetics terms.

Additional information

Funding

Funder	Grant reference number	Author
National Key Research and Development Program of China	2021ZD0201501	Yi Du
National Natural Science Foundation of China	31822024	Yi Du
Strategic Priority Research Program of Chinese Academy of Sciences	XDB32010300	Yi Du
The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.		

Author contributions

Conceptualization, Y.D.; Methodology, B.S.L., W.Y.Z. and Y.D.; Software, B.S.L., and W.Y.Z.; Investigation, B.S.L., and Y.C.L.; Formal Analysis, B.S.L., Y.C.L.; Visualization: B.S.L.; Writing – Original Draft, B.S.L.; Writing – Review & Editing, Y.D.; Funding Acquisition: Y.D.; Supervision: Y.D.

Author ORCIDs

Baishen Liang http://orcid.org/0000-0003-2604-9120
Yanchang Li http://orcid.org/0000-0003-1741-8596
Wanying Zhao http://orcid.org/0000-0002-6979-940X

Yi Du http://orcid.org/0000-0003-4512-5221

Resources table

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Matlab R2016a & R2021a	MathWorks	RRID: SCR_001622
Psychtoolbox-3	http://psychtoolbox.org/	RRID: SCR_002881
Praat 6	http://www.fon.hum.uva.nl/praat/	RRID: SCR_016564
G*Power	Heinrich Heine University Dusseldorf	RRID: SCR_013726
Brainsight	Rogue Research	RRID: SCR_009539
MRIcron	https://www.nitrc.org/projects/mricron/	RRID: SCR_002403
Analysis of Functional NeuroImages	https://afni.nimh.nih.gov/	RRID: SCR_005927
BrainNet Viewer 1.7	https://www.nitrc.org/projects/bnv/	RRID: SCR_009446
Legacy STRAIGHT	https://github.com/ HidekiKawahara/legacy_STRAIGHT	RRID: SCR_021957

Declaration of interests

The authors declare no competing interests.

Additional files

Supplemental files

Audio S1. Clear syllables [ti55], [ti35], [t^hi55], and [t^hi35] for continua synthesis.

Audio S2. Morphed five-step lexical tone (F0) continuum from [ti55] to [ti35].

Audio S3. Morphed five-step consonant (VOT) continuum from [ti55] to [t^hi55].

Audio S2–S3 were selected from the individualized stimuli matrix from one participant in Experiment 1.

Data and code availability

Raw data and code for analyses in this study will be deposited in the Institutional Repository, Institute of Psychology, Chinese Academy of Sciences (http://ir.psych.ac.cn/) after publication. Custom code for the experiment, and visualization are available by request to the corresponding author.

References

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*, 1043–1047. https://doi.org/10.1126/science.aaz3468

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), 493–498. https://doi.org/10.1038/s41586-019-1119-1

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex*, *13*(1), 85–95. https://doi.org/10.1016/S0010-9452(77)80057-9

Behroozmand, R., Phillip, L., Johari, K., Bonilha, L., Rorden, C., Hickok, G., & Fridriksson, J. (2018). Sensorimotor impairment of speech auditory feedback processing in aphasia. *NeuroImage*, *165*, 102–111. https://doi.org/10.1016/j.neuroimage.2017.10.014

Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience and Biobehavioral Reviews*, *77*, 177–193. https://doi.org/10.1016/j.neubiorev.2017.03.014

Bergmann, T. O., & Hartwigsen, G. (2020). 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

Beynel, L., Appelbaum, L. G., Luber, B., Crowell, C. A., Hilbig, S. A., Lim, W., Nguyen, D., Chrapliwy, N. A., Davis, S. W., Cabeza, R., Lisanby, S. H., & Deng, Z. De. (2019). Effects of online repetitive transcranial magnetic stimulation (rTMS) on cognitive processing: a meta-analysis and recommendations for future studies. *Neuroscience and Biobehavioral Reviews*, *107*, 47–58. https://doi.org/10.1016/j.neubiorev.2019.08.018

Bloom, J. S., & Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychology Review*, *15*(2), 59–71. https://doi.org/10.1007/s11065-005-6252-y

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, *15*, 371–383. https://doi.org/10.1016/0028-3932(77)90089-6

Bolgnini, N., & Ro, T. (2010). Transcranial magnetic stimulation: disrupting neural activity to alter and assess brain function. *Journal of Neuroscience*, *30*(29), 9647–9650. https://doi.org/10.1523/JNEUROSCI.1990-10.2010

Brodbeck, C., Bhattasali, S., Cruz Heredia, A. AL, Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *ELife*, *10*, e72056. https://doi.org/10.7554/elife.72056

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251. https://doi.org/10.1017/S0952675700001019

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*, 837–845. https://doi.org/10.1093/cercor/bhm131

Chao, Y. R. (1968). *Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chao, Y. R. (1980). A system of “tone-letters.” *Fangyan[Dialet]*, *2*, 81–83.

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*, *41*(7), 1842–1858. https://doi.org/10.1002/hbm.24916

Coffey, E. B. J., Herholz, S. C., Scala, S., & Zatorre, R. J. (2011). Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research. *The Neurosciences and Music IV: Learning and Memory*.

D’Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, *118*, 9–14. https://doi.org/10.1016/j.bandl.2011.02.007

D’Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2011). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, *48*, 882–887. https://doi.org/10.1016/j.cortex.2011.05.017

D’Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *369*, 20130418. https://doi.org/10.1098/rstb.2013.0418

D’Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385. https://doi.org/10.1016/j.cub.2009.01.017

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The control of vocal pitch in human laryngeal motor cortex. *Cell*, *174*, 21–31. https://doi.org/10.1016/j.cell.2018.05.016

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*, 78–89. https://doi.org/10.1152/jn.00297.2011

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(19), 7126–7131. https://doi.org/10.1073/pnas.1318738111

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications*, *7*, 12241. https://doi.org/10.1038/ncomms12241

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(51), 13579–13854. https://doi.org/10.1073/pnas.1712223114

Eichert, N., Papp, D., Mars, R. B., & Watkins, K. E. (2020). Mapping human laryngeal motor cortex during vocalization. *Cerebral Cortex*, *30*, 6254–6269. https://doi.org/10.1093/cercor/bhaa182

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*, 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology*, *117*(12), 2584–2596. https://doi.org/10.1016/j.clinph.2006.06.712

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, *3*(4), 393–405. https://doi.org/10.1038/s41562-019-0548-z

Friston, K. J., & Price, C. J. (2003). Degeneracy and redundancy in cognitive anatomy. *Trends in Cognitive Sciences*, *7*(4), 141–152. https://doi.org/10.1016/S1364-6613(03)00034-2

Gandour, J., Ponglorpisit, S., & Dardarananda, R. (1992). Tonal disturbances in Thai after brain damage. *Journal of Neurolinguistics*, *7*, 133–145. https://doi.org/10.1016/0911-6044(92)90015-O

Ge, J., Peng, G., Lyu, B., Wang, Y., Zhuo, Y., Niu, Z., Tan, L. H., Leff, A. P., & Gao, J.-H. (2015). Cross-language differences in the brain network subserving intelligible speech. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 2972–2977. https://doi.org/10.1073/pnas.1416000112

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics*. Wiley-Blackwell.

Güntürkün, O., Ströckens, F., & Ocklenburg, S. (2020). Brain lateralization: a comparative perspective. *Physiological Reviews*, *100*(3), 1019–1063. https://doi.org/10.1152/physrev.00006.2019

Hartwigsen, G., Weigel, A., Schuschan, P., Siebner, H. R., Weise, D., Classen, J., & Saur, D. (2016). Dissociating parieto-frontal networks for phonological and semantic word

decisions: a condition-and-perturb TMS study. *Cerebral Cortex*, *26*(6), 2590–2601. https://doi.org/10.1093/cercor/bhv092

Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to Wilson: what does motor cortex contribute to speech perception? *Trends in Cognitive Sciences*, *13*(8), 330–331. https://doi.org/10.1016/j.tics.2009.05.002

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

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

Hincapié Casas, A. S., Lajnef, T., Pascarella, A., Guiraud-Vinatea, H., Laaksonen, H., Bayle, D., Jerbi, K., & Boulenger, V. (2021). Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. *NeuroImage*, *244*, 118577. https://doi.org/10.1016/j.neuroimage.2021.118577

Huang, B., & Liao, X. (2017). *Xiandai Hanyu[Modern Chinese]* (6th ed.). Higher Education Press.

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. https://doi.org/10.1126/science.aax0287

Keitel, A., Ince, R. A. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment

interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, *147*, 32–42. https://doi.org/10.1016/j.neuroimage.2016.11.062

Ladefoged, P., & Johnson, K. (2010). *A Course in Phonetics* (J. M. Flaherty (ed.); 6th ed.). Michael Rosenberg.

Lévêque, Y., Muggleton, N., Stewart, L., & Schön, D. (2013). Involvement of the larynx motor area in singing voice perception: a TMS study. *Frontiers in Psychology*, *4*, 418. https://doi.org/10.3389/fpsyg.2013.00418

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: an activation likelihood estimation meta-analysis. *Frontiers in Neuroscience*, *12*, 495. https://doi.org/https://doi.org/10.3389/fnins.2018.00495

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Mars, R. B., Eichert, N., Jbabdi, S., Verhagen, L., & Rushworth, M. F. S. (2018). Connectivity and the search for specializations in the language-capable brain. *Current Opinion in Behavioral Sciences*, *21*, 19–26. https://doi.org/10.1016/j.cobeha.2017.11.001

Mas-herrero, E., Dagher, A., & Zatorre, R. J. (2018). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. *Nature Human Behaviour*, *2*, 27–32. https://doi.org/10.1038/s41562-017-0241-z

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension:

state of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, *23*, 1190–1197. https://doi.org/10.1093/cercor/bhs110

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience*, *29*(31), 9819–9825. https://doi.org/10.1523/JNEUROSCI.6018-08.2009

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia*, *94*, 13–22. https://doi.org/10.1016/j.neuropsychologia.2016.11.016

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mouncastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*, 1256846. https://doi.org/10.1126/science.1256846

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: a theory and

review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

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

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*, 62–66. https://doi.org/10.1126/science.aax0050

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

Sammler, D., Grosbras, M.-H. H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*(23), 1–7. https://doi.org/10.1016/j.cub.2015.10.009

Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neuroscience and Biobehavioral Reviews*, *35*(3), 516–536. https://doi.org/10.1016/j.neubiorev.2010.06.005

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS

evidence. *Cerebral Cortex*, *25*, 3894–3902. https://doi.org/10.1093/cercor/bhu274

Si, X., Zhou, W., & Hong, B. (2017). Cooperative cortical network for categorical processing of Chinese lexical tone. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), 12303–12308. https://doi.org/10.1073/pnas.1710752114

Simonyan, K. (2014). The laryngeal motor cortex: its organization and connectivity. *Current Opinion in Neurobiology*, *28*, 15–21. https://doi.org/10.1016/j.conb.2014.05.006

Suppa, A., Huang, Y. Z., Funke, K., Ridding, M. C., Cheeran, B., Di Lazzaro, V., Ziemann, U., & Rothwell, J. C. (2016). Ten years of theta burst stimulation in humans: established knowledge, unknowns and prospects. *Brain Stimulation*, *9*(3), 323–335. https://doi.org/10.1016/j.brs.2016.01.006

Tamura, S., Hirose, N., Mitsudo, T., Hoaki, N., Nakamura, I., Onitsuka, T., & Hirano, Y. (2022). Multi-modal imaging of the auditory-larynx motor network for voicing perception. *NeuroImage*, *251*, 118981. https://doi.org/10.1016/j.neuroimage.2022.118981

Tang, D. L., Möttönen, R., Asaridou, S. S., & Watkins, K. E. (2021). Asymmetry of auditory-motor speech processing is determined by language experience. *Journal of Neuroscience*, *41*(5), 1059–1067. https://doi.org/10.1523/JNEUROSCI.1977-20.2020

Thickbroom, G. W. (2007). Transcranial magnetic stimulation and synaptic plasticity: experimental framework and human models. *Experimental Brain Research*, *180*(4), 583–593. https://doi.org/10.1007/s00221-007-0991-3

Thut, G., Veniero, D., Romei, V., Miniussi, C., Schyns, P., & Gross, J. (2011). Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Current Biology*, *21*(14), 1176–1185. https://doi.org/10.1016/j.cub.2011.05.049

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews Neuroscience*, *4*(1), 37–48. https://doi.org/10.1038/nrn1009

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience*, *16*, 978–987. https://doi.org/10.1162/0898929041502616

Wilson, M. (2002). Six views of embodied cognition. *Psychometric Bulletin & Review*, *9*(4), 625–636.

Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences*, *13*(8), 329–330. https://doi.org/10.1016/j.tics.2009.06.001

Witteman, J., Van Ijzendoorn, M. H., Van de Velde, D., Van Heuven, V. J. J. P., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia*, *49*, 3722–3738. https://doi.org/10.1016/j.neuropsychologia.2011.09.028

Wu, X., Wang, C., Chen, J., Qu, H., Li, W., Wu, Y., Schneider, B. A., & Li, L. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research*, *199*, 1–10. https://doi.org/10.1016/j.heares.2004.03.010

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849. https://doi.org/10.1126/science.1589767

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

Ziemann, U. (2010). TMS in cognitive neuroscience: virtual lesion and beyond. *Cortex*, *46*(1), 124–127. https://doi.org/10.1016/j.cortex.2009.02.020

Appendix

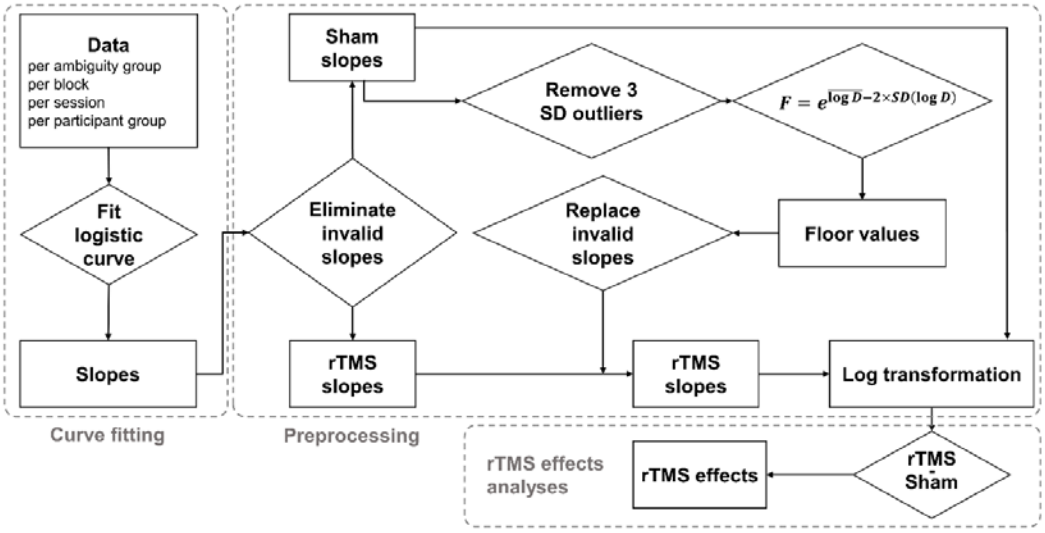
Acoustic properties of stimuli

Appendix Table 1. Acoustic properties of clear syllables, and individualized synthesized continua used in Experiment 1 and 2. Note that for Experiment 1, [ti^{h}55] – [ti^{h}35] and [ti35] – [ti^{h}35] were not shown. Values depict mean ± standard deviation.

VOT: voice onset time. F0 was extracted by the *pitch* function in MatlabR2021a. F0 glide was calculated by subtracting pitch onsets from pitch offsets.

Stimuli	VOT (ms)	Duration (ms)	mean F0 (Hz)	F0 glide (Hz)	F0 glide (Hz/s)
Clear syllables					
[ti55]	0.00	355.94	334.63	7.65	27.33
[ti35]	0.00	354.06	271.39	105.75	377.68
[ti^{h}55]	88.68	444.63	334.63	7.65	27.33
[ti^{h}35]	88.68	442.74	271.39	105.75	377.68
Experiment 1					
[ti55]	3.41 ± 0.00	357.39 ± 0.00	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
	3.41 ± 0.00	357.39 ± 0.00	327.53 ± 2.74	16.01 ± 4.16	53.37 ± 13.87
|	3.41 ± 0.00	357.39 ± 0.00	321.95 ± 5.46	24.81 ± 8.28	82.70 ± 27.59
	3.41 ± 0.00	357.39 ± 0.00	316.26 ± 8.08	33.10 ± 12.17	110.34 ± 40.55
[ti35]	3.41 ± 0.00	357.39 ± 0.00	310.63 ± 10.81	41.43 ± 15.90	138.10 ± 53.00
[ti55]	3.41 ± 0.00	357.39 ± 0.00	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
	16.05 ± 2.93	370.02 ± 2.93	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
|	28.68 ± 5.86	382.66 ± 5.86	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
	41.32 ± 8.79	395.31 ± 8.79	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
[ti^{h}55]	53.96 ± 11.72	407.93 ± 11.72	333.22 ± 0.06	7.62 ± 0.63	25.40 ± 2.12
Experiment 2					
[ti55]	1.54 ± 1.60	355.52 ± 1.60	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99
	1.54 ± 1.60	355.52 ± 1.60	327.63 ± 1.01	16.38 ± 1.57	54.61 ± 5.23
|	1.54 ± 1.60	355.52 ± 1.60	321.72 ± 2.04	25.29 ± 3.15	84.29 ± 10.48
	1.54 ± 1.60	355.52 ± 1.60	315.82 ± 3.11	34.11 ± 4.86	113.70 ± 16.19
[ti35]	1.54 ± 1.60	355.52 ± 1.60	309.93 ± 4.07	43.10 ± 6.39	143.67 ± 21.31
[ti55]	1.54 ± 1.60	355.52 ± 1.60	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99
	11.36 ± 3.10	365.24 ± 3.13	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99
|	21.17 ± 5.49	375.13 ± 5.53	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99
	30.99 ± 8.00	384.95 ± 8.04	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99
[ti^{h}55]	40.80 ± 10.56	394.78 ± 10.55	333.58 ± 0.58	7.16 ± 0.90	23.88 ± 2.99

Behavioral data analysis



Appendix Figure 1. Procedure of curve fitting, preprocessing of slopes, and rTMS effects analyses for Experiment 1 (Those for Experiment 2 were similar)

Competitions between consonant and tone perception

To explore interactions between consonant and tone perception, we used data from sham conditions in Experiment 1 (Experiment 1 Sham) as interference from TMS was excluded. Participants in left and right hemisphere stimulation groups were pooled together. Here, to extract slopes, for each block of tasks, trials were categorized into ambiguous (step 3), half-ambiguous (step 2 and 4), and unambiguous (step 1 and 5) conditions according to the unattended dimension of a matrix (Appendix figure 2A). This allows us to test the main effects of ambiguity of the unattended feature as well as to quantify the task difficulty as a function of such ambiguity.

To confirm the reliability of potential effects, we also analyzed data from two other unpublished tests (Pretest 1 & 2) conducted previously. Both tests used paradigms similar to this study, but in groups of participants that did not enter the current study.

22 right-handed young adults (13 females, mean age 21.32 years, SD = 2.16)

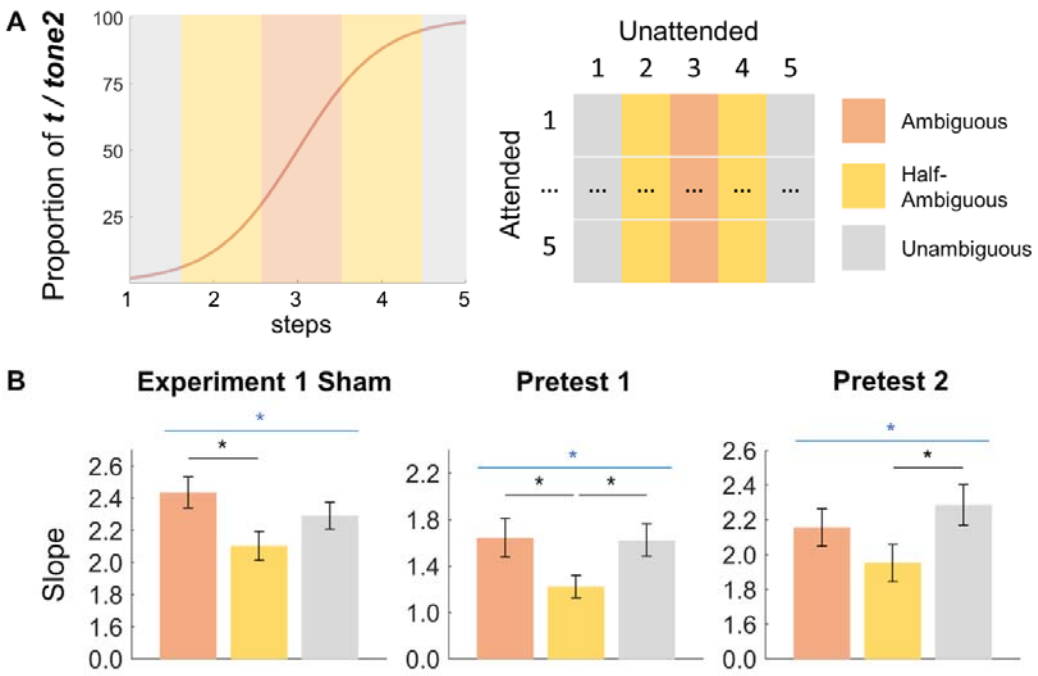
participated in Pretest 1. The group size is sufficient to detect medium to large effects as estimated by G*Power3. They had normal hearing and gave written informed consent prior to the experiment. They were paid ￥50 per hour for participation. The experiment was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences. Participants performed speeded tone/consonant categorical identification tasks with/without noise similar to those in Experiment 1. Stimuli came from a 7×7-steps tone (high-level tone to mid-rising tone) – consonant ([t] to [t^h]) matrix of continua. One token was presented in each trial and each block contained 200 trials (more trials around PSE to guarantee curve fitting), to which different tokens were randomly assigned. Inter-trial interval (ITI) was on average 0.5s (0.4 – 0.6s, 10 ms step, uniformly distributed). For each block, trials were categorized into ambiguous (step 4), half-ambiguous (step 2, 3, 5, and 6), and unambiguous (step 1 and 7) conditions according to the unattended dimension of a matrix.

58 right-handed young adults (29 females, mean age = 21.31 years, SD = 2.93) participated in Pretest 2. The group size is sufficient to detect medium to large effects as estimated by G*Power3. They had normal hearing and gave written informed consent prior to the experiment. They were paid ￥50 per hour for participation. The experiment was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences. Participants performed speeded tone and consonant categorical identification tasks similar to those in Experiment 1, but only in quiet. Stimuli came from a 5×5-step matrix of continua (high-level tone to mid-rising tone)–([t] to [t^h]). Each block contained 39 trials, to which different tokens were randomly assigned. ITI was identical to that in Pretest 1. For each block, trials were categorized into ambiguous (step 3), half-ambiguous (step 2 and 4), and unambiguous (step 1 and 5) conditions according to the unattended dimension of a matrix.

The fitting and preprocessing procedures for Experiment 1 Sham, Pretest 1, and Pretest 2 were identical to those in TMS experiments (**Appendix Figure 1**). Notably, curve fitting was performed separately for each ambiguity condition of trials, and block-level upper and lower asymptotes were calculated by unconstrained fit on data.

Slopes of psychometric curves in each block were pooled by averaging them across blocks (tone or consonant blocks for all three tests, and with/without noise mask for Experiment 1 Sham and Pretest 1) separately for ambiguous, half-ambiguous, and unambiguous conditions of trials. For each test, the main effects of ambiguity of the unattended feature on the slopes were tested. To do so, we applied one-way repeated measures analysis of variance (rANOVA) to Experiment 1 Sham and Pretest 1. For Pretest 2, however, since Kolmogorov-Smirnov test of normality showed that slopes in the unambiguous condition were not normally distributed, Friedman test, a non-parametric alternative of rANOVA was used instead. Results showed that, main effects of ambiguity of the unattended feature were significant in all three tests (Experiment 1 Sham: $F_{2,74}$ = 4.266, p = 0.018, partial η^2 = 0.103; Pretest 1: $F_{2,30}$ = 5.001, p = 0.013, partial η^2 = 0.250; Pretest 2: $\chi^2(2)$ = 6.259, p = 0.044, Kendall's W = 0.058).

Moreover, we applied post-hoc contrast analyses to investigate how the ambiguity of the unattended feature affects tone and consonant categorical perception. Results showed that, for all three tests, slopes were lowered when the unattended feature was half-ambiguous. In Experiment 1 Sham, slopes from the half-ambiguous condition were lower than those from the ambiguous condition (Mean difference = -0.331, p = 0.022, Bonferroni corrected). In Pretest 1, slopes from the half-ambiguous condition were lower than those from the ambiguous (Mean difference = -0.425, p = 0.042, Bonferroni corrected) and unambiguous (Mean difference = -0.403, p = 0.021, Bonferroni corrected) conditions. In Pretest 2, slopes from the half-ambiguous condition were lower than those from the unambiguous condition (Z = -2.820, p = 0.014, Bonferroni corrected, Wilcoxon signed-rank test). Presumably, processing of the unattended feature of a syllable may occupy neural resources for the parallel computations of the attended feature (Luber & Lisanby, 2014), and such competition was amplified when the unattended feature was half-ambiguous as participants had biased but uncertain perception.



Appendix figure 2. Competitions between tone and consonant perception

(**A**) Take Experiment 1 Sham as an example: trials were grouped according to the ambiguity of the unattended feature. Left: the ambiguity is a function of stimuli continuum (participants made confident judgements at unambiguous steps, biased but uncertain judgements at half-ambiguous steps, and guessing judgements at ambiguous steps). Right: trials were grouped into 3 conditions according to unambiguous, half ambiguous, and ambiguous steps in the unattended dimension. "t" represents [t^h]; "tone2" represents mid-rising tone. Orange: ambiguous conditions; yellow: half-ambiguous conditions; grey: unambiguous conditions.

(**B**) The ambiguity of the unattended feature affected categorical speech perception. The main effects of ambiguity on slopes of psychometric curves were significant in Experiment 1 Sham (left), Pretest 1 (middle), and Pretest 2 (right). Post-hoc analyses showed that slopes were significantly lowered in half-ambiguous conditions compared with ambiguous (Experiment 1 Sham and Pretest 1) and unambiguous (Pretest 1 and Pretest 2) conditions. Blue horizontal lines and asterisks represent significant main effects of ambiguity; black horizontal lines and asterisks represent significant differences between ambiguity conditions. Statistical tests were performed by rANOVA for Experiment 1 Sham and Pretest 1, and by non-parametric Friedman test and Wilcoxon signed ranks test for Pretest 2. P values for post hoc analyses were adjusted by Bonferroni's multiple comparison corrections. * corrected $p < 0.05$.

Functional localization experiment

Procedure

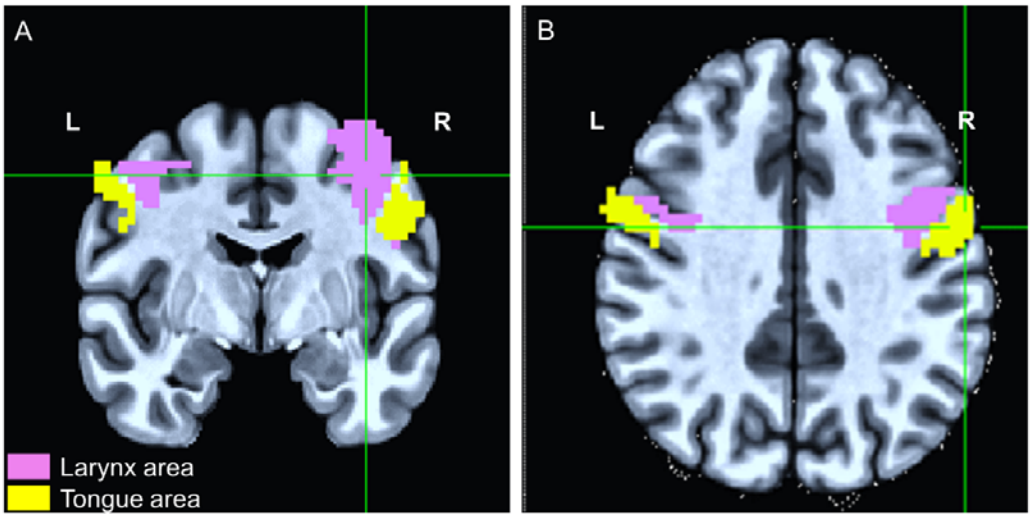
48 right-handed participants (26 females, mean age = 21.44, SD = 2.79) took part in the functional localization experiment of speech motor areas (i.e., LMC and TMC). The group size is sufficient to detect medium to large effects as estimated by G*Power3. They gave written informed consent prior to the experiment. They were paid ￥100 per hour for participation. The experiment was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences. 10 participants also took part in experiment 1 (N = 9) and experiment 2 (N = 1). Before the experiment, T1-weighted anatomical images were acquired on a 3-Tesla Siemens Magnetom Trio scanner using a 20-channel head coil, using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2200 ms, TE = 3.49 ms, slices per slab = 192, field of view FOV = 256 mm, flip angle = 8°, spatial resolution = 1×1×1 mm), identical to experiment 1 and 2. During the experiment, participants needed to articulate the voiceless [t^h] (only tongue moved) or pronounce the voiced [a] (only laryngeal vocal fold vibrated), to localize the TMC and the LMC, respectively. The experiment was in a block design and was divided into 8 blocks, including 4 TMC and 4 LMC localization blocks. The order of blocks was counterbalanced across participants. A resting period with the same duration followed each block. Each block contained 8 trials, each lasting for 2s. To avoid breathing artifacts, before the beginning of each block, participants took a deep breath and held their breath until the block ended. In each trial, participants articulated or pronounced sounds following the visual cue (“Say AH” or “Say D”) projected to the screen. Functional images were acquired by a continuous multiband-accelerated echo-planar imaging sequence (multiband factor = 4, 40 slices, TR = 640 ms, TE = 30 ms, flip angle = 25°, FOV = 192 mm, voxel size = 3 × 3 × 3 mm). Blood oxygen level-dependent (BOLD) signals were preprocessed and analyzed at individual and group levels.

fMRI data analysis

Functional images were processed by the AFNI package (Cox, 1996), including slice-timing correction, physiological motion correction, realignment to an MNI reference image, and spatial smoothing (Gaussian filter FWHM = 6.0 mm).

Generalized linear model (GLM) analyses were conducted at the individual level. The predicted time course of BOLD activation was modeled as a “box-car” function convolved with the canonical hemodynamic response function. Individual contrast maps were defined by comparing “AH” blocks to “D” blocks (LMC: “AH” – “D”, TMC: “D” – “AH”) using paired t-test. Contrast maps were then subjected to group analysis where one-sample t-test (alternative hypothesis $\mu = 0$) was used to find activated brain areas, which were masked by ROIs of bilateral precentral gyrus in automated anatomical labeling (AAL) templates (Tzourio-Mazoyer et al., 2002).

MNI coordinates of bilateral homologous LMC and TMC were determined by visual inspection on the group map of activation. Criteria of target selection included: (1) bilateral homologous targets should share a similar degree of activation; (2) geometric middle points, instead of peaks of activation areas, were selected to ensure that magnetic stimulation covers the maximal volume of activations. The chosen MNI coordinates were: [±40, -5, 50] for LMC and [±59, -3, 36] for TMC. Note that, although the two selected targets were proximal to each other, their spatial distance still exceeds the effective range of a figure-eight TMS coil (Bolgnini & Ro, 2010). Thus, transmission effects of stimulation can be ignored.



Appendix figure 3 : Group-level activation in the functional localization experiment.
Activation areas were masked by AAL ROIs of the bilateral precentral gyrus. (A) Dorsal-ventral functional dissociation between larynx (pink) and tongue (yellow) motor representation was shown in the coronal view of activation, and the green crosshair was focused on the LMC. (B) The axial view showed the position of TMC, as was focused by the green crosshair. L, left hemisphere; R, right hemisphere.

Appendix references

Bolgnini, N., & Ro, T. (2010). Transcranial magnetic stimulation: disrupting neural activity to alter and assess brain function. *Journal of Neuroscience*, *30*(29), 9647–9650. https://doi.org/10.1523/JNEUROSCI.1990-10.2010

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

Luber, B., & Lisanby, S. H. (2014). Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *NeuroImage*, *85*, 961–970. https://doi.org/10.1016/j.neuroimage.2013.06.007

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978