A causal role of human laryngeal motor cortex in perceptual decision of lexical pitch and voicing

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

Speech motor cortices are believed to simulate articulatory features of the upcoming speech to aid perception, especially under adverse listening conditions. However, whether and how the laryngeal motor cortex (LMC), which is essential for regulating linguistic pitch and voicing during production, is engaged in speech perception remains elusive. Here, we applied transcranial magnetic stimulation (TMS) upon the left or right dorsal LMC (dLMC) localized by a neuroimaging articulation test, and participants performed perceptual decision tasks on Mandarin lexical tone or consonant (dental plosive) presented with/without noise. Using psychometric function and hierarchical drift-diffusion modeling to disentangle perceptual sensitivity and decision-making parameters, we found a causal role of dLMC in the perceptual decision of both lexical pitch and voicing with a left-lateralized engagement. Moreover, task difficulty modulated the involvement as medium-level demands optimize its contributions. Our findings provide new causal evidence for region-specific and modulatory motor engagement in speech perception and decision.

Teaser

Laryngeal motor cortex engages in the perceptual decision of lexical pitch and voicing, as they does in speech production.

Keywords:

laryngeal motor cortex, speech perception, perceptual decision, lexical tone, voicing

Introduction

Human speech is characterized by articulatory features determined by motor gestures of the vocal folds and the vocal tract[1], ultimately controlled by the speech motor system in the brain[2]. Intriguingly, speech perception has been postulated to recruit motoric simulation by the cortical speech motor system. The classical motor theory of speech perception proposes that recognition of phonemes is achieved by reconstructing their articulatory motor sequences[3]. In the dual-stream models of speech processing that emphasize a reciprocal auditory-motor transformation mechanism[4–6], the articulatory network in the dorsal stream maps phonological analyses onto motor representations during speech perception.

Indeed, recent transcranial magnetic stimulation (TMS) studies have revealed a causal role of the motor cortex in the perceptual decision of phonemes in an effector-specific somatotopic manner. A lip motor subregion is engaged in perceiving bilabial consonants[7–9], whereas a tongue motor area is recruited in distinguishing dental consonants[9–11] and vowels[12]. However, compared with the tongue and lip motor areas, the contribution of the laryngeal motor cortex (LMC) to speech perception has rarely been investigated. Physiologically, the LMC, especially the dorsal LMC (dLMC) regulates vocal pitch contour and controls voicing during speech and singing[13,14] by altering the tension and adduction/abduction of the vocal folds[15] in the larynx. This allows speakers to produce pitch contour (e.g., Mandarin lexical tone and speech prosody) and voicing (e.g., voice onset time, VOT) contrasts, respectively. Therefore, we hypothesized that the dLMC would be causally involved in the perceptual decision of phonemic contrasts characterized by pitch contour and VOT, mirroring its indispensability in speech production (see Fig. 1A). We also hypothesized that the recruitment of the dLMC would be modulated by task demands since previous neuroimaging studies have demonstrated that the motor engagement can constrain and compensate for degraded speech perception as listening context requires[16–19] (for a review see[20]).

Here, we delivered repetitive TMS (rTMS, Experiment 1) or theta-burst

stimulation (TBS, Experiment 2) to Mandarin speakers' left or right dLMC to investigate if the identification performance of lexical tone and consonant with VOT contrast in quiet or in noisy background would be modulated accordingly. If the dLMC is essential in pitch contour and voicing categorization, we would expect that TMS upon the dLMC alters the performance (impairs perceptual sensitivity as reflected by a decreased slope of identification curve and vice versa) in both tone and consonant identification tasks (Fig. 1a and Fig. 1d). We further applied the drift-diffusion model (DDM)[21] to single-trial binary responses and reaction times (RTs) in Experiment 2 to disentangle what latent psychological processes underlying perceptual decision-making the dLMC is engaged in (Fig. 1e). Moreover, by comparing TMS effects in the left and right dLMC stimulation conditions, and in tasks with quiet and noisy background, we aimed at clarifying the hemispheric asymmetry and regulations by task difficulty of the dLMC engagement.

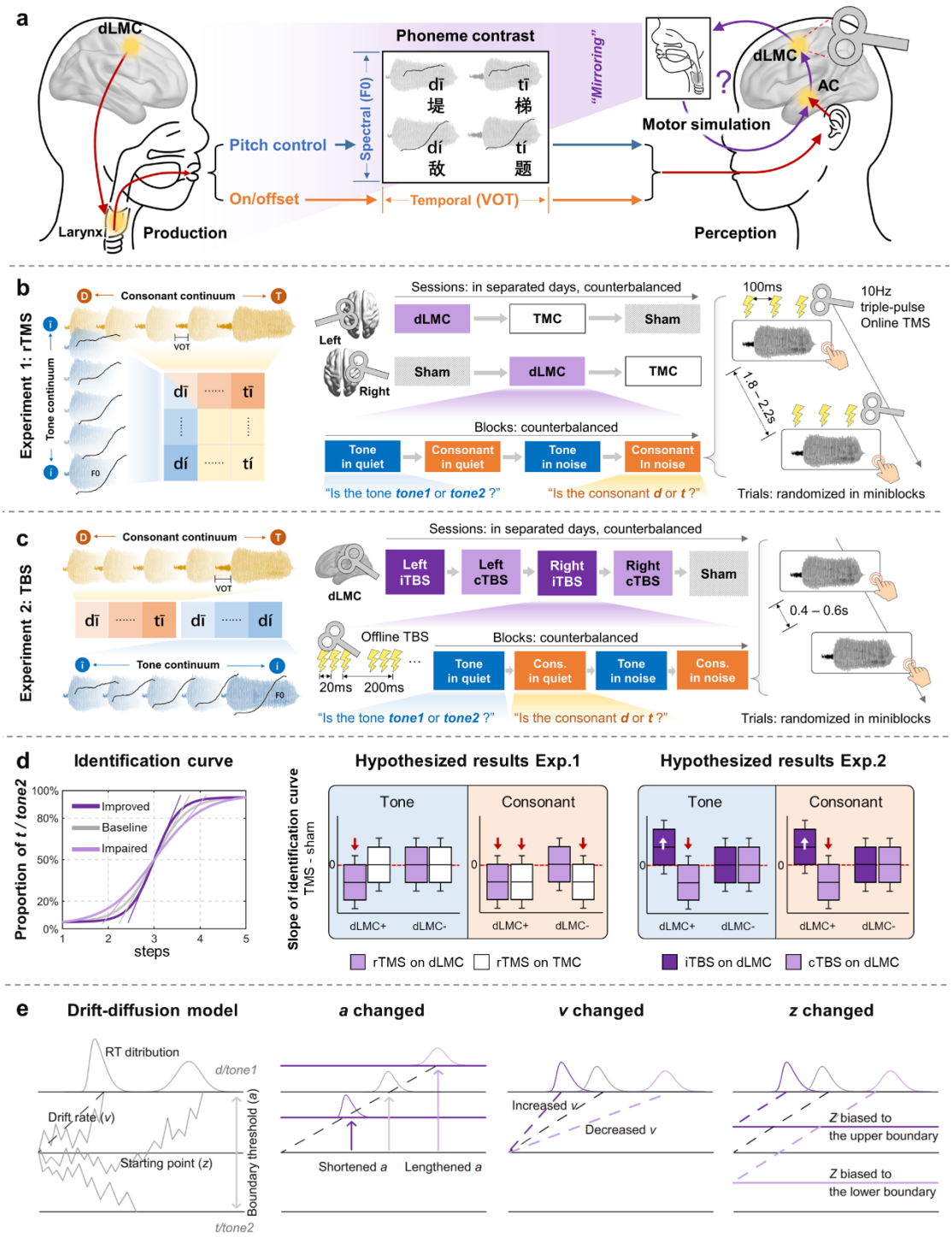

Figure 1. Experimental design and hypothesized results.

(**a**) Schematic illustration of sensorimotor integration in Mandarin speech perception and protocol of this study. The dLMC commands the larynx to control the pitch of voice and the onset and offset of phonation during speech articulation. Correspondingly, the dLMC is assumed to engage in the 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[22], 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 dLMC before the task.

(**d**) Hypothetical results for psychometric curve fitting. 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 dLMC is engaged in Mandarin speech perception, compared with sham baseline (red dashed lines). dLMC+/-: when the dLMC 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.

(**e**) Hypothetical results for the drift-diffusion model (DDM). The DDM (left) regards the two-alternative perceptual decision as an evidence accumulation process in noise. Evidence accumulates from a starting point (z, also represents response bias) with a drift rate (v) until reaching one of the boundaries (with a threshold a). The DDM predicts distributions of response accuracies and reaction times. The middle to right three panels show hypothetical models with each of the three parameters changed by TMS upon the dLMC. This demonstration is adopted from Ratcliff et al. (2016)[23]. Note that, non-decision processes are not shown.

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 Fig. 2 and Fig. 3.

dLMC: dorsal laryngeal motor cortex; TMC: tongue motor cortex.

Results

The dLMC is causally involved in the perception of lexical pitch and voicing

Experiment 1

In Experiment 1, we used rTMS to estimate changes in Mandarin lexical tone and consonant perception in a group of 64 young adults. Participants were separated into two matched groups for left and right dLMC stimulation, respectively. Experiment 1 included 3 rTMS sessions: sham, dLMC stimulation, and 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[9–11]. The MNI coordinates of the dLMC [±40, -5, 50] and the TMC [±59, -3, 36] were defined by a functional magnetic resonance imaging (fMRI) localization task in 48 Mandarin speakers (see Supplementary Information, Functional localization experiment). Note that, we only found activations in the dorsal but not ventral LMC. Within each session, participants underwent four blocks of syllable identification tasks: tone and consonant tasks with or without noise masking where they performed force choice identification judgments. Syllables were randomly selected from a 5×5-step tone–consonant continuum matrix (Fig. 1b, left panel). Began with the syllable onsets, 3 pulses of 10-Hz biphasic rTMS were applied to the dLMC or the TMC.

Slopes of psychometric functions served as the response variable. The lower the slope, the weaker the participant's capacity to identify phonemes (Fig. 1d, left panel). In other words, the slope quantifies perceptual sensitivity that is crucial to phoneme categorization[24]. Unexpectedly, we found competitions between tone and consonant perception such that slopes were affected by the ambiguity of the unattended dimension (see Supplementary Information, 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 (see Methods, Experiment 1, Stimulation effect analyses).

Results showed that rTMS upon the dLMC exerted inhibitory effects on both tone and consonant perception. Tone perception in noise masking was impaired by

stimulating the left dLMC 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, Fig. 2c), and by stimulating the right dLMC 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, Fig. 2b). Left dLMC stimulation also impaired consonant perception without noise-masking when tones were unambiguous (compared with sham: dLMC, $p_{fdr} = 0.033$, Cohen's d = -0.459, permutation test, Fig. 2a). Meanwhile, consistent with previous studies[9–11], 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, Fig. 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 dLMC.

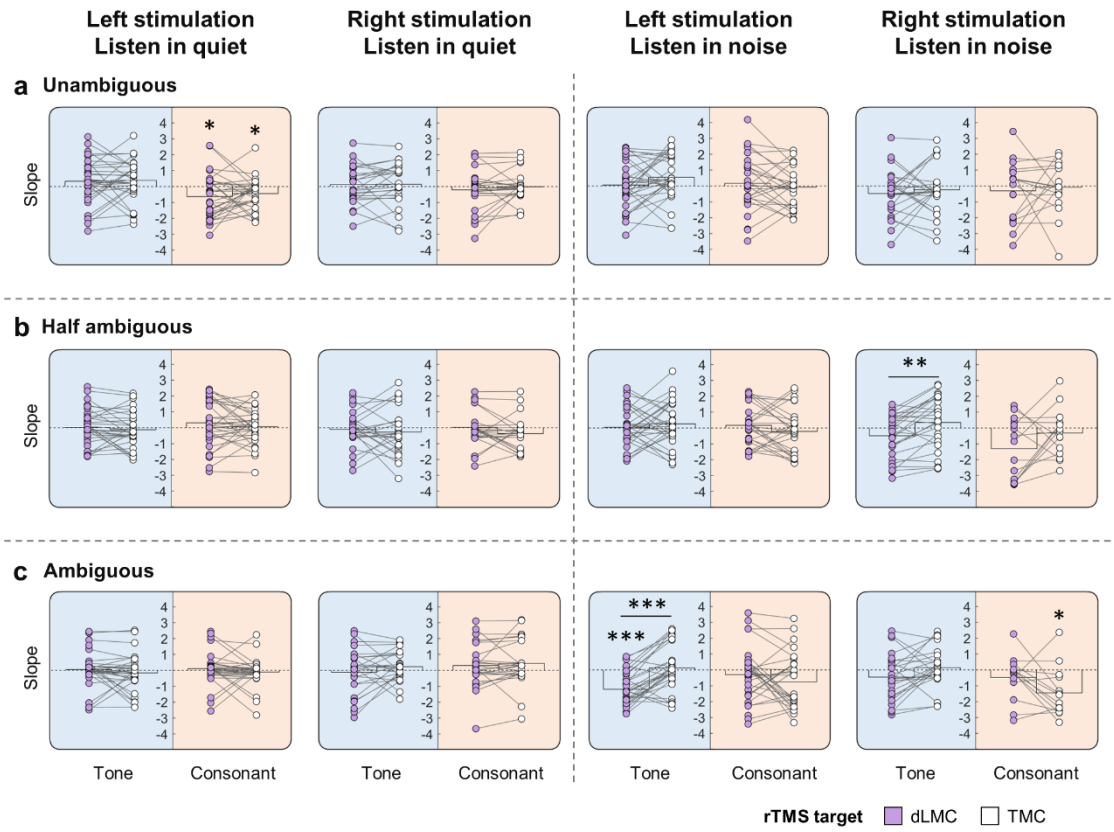

Figure 2. rTMS effects on slopes of tone and consonant identification curves in Experiment 1.

(**a**) When syllables were unambiguous in the unattended dimension, rTMS upon the left dLMC 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 dLMC impaired tone perception in noise.

(**c**) When syllables were ambiguous in the unattended dimension, rTMS upon the left dLMC 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 the dLMC 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.

dLMC: dorsal 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 dLMC with TMC within the same tasks. P values (single-tailed) were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

Experiment 2

Although Experiment 1 found some evidence for the dLMC engagement in lexical tone and consonant perception, results were inconsistent. And as we performed curve-fittings after separating trials into three groups of ambiguity, the reliability of slope estimations was questionable due to limited trial numbers[24]. To replicate findings of dLMC stimulation in Experiment 1, we used offline TBS in another group of 26 participants in Experiment 2. This is a TMS paradigm whose directionality has been extensively explored as intermittent (iTBS) and continuous (cTBS) stimulation would increase and decrease cortical excitability, respectively[25]. Moreover, to avoid competitions between tone and consonant, in Experiment 2 we used one F0 and one VOT continuum, instead of a matrix, for the tone and consonant task, respectively (Fig. 1c, left). Furthermore, to directly estimate hemispheric asymmetry of the dLMC in speech perception, each participant received TBS upon both hemispheres in five separated sessions in a random order: sham, iTBS on the left dLMC, cTBS on the left dLMC, iTBS on the right dLMC, and cTBS on the right dLMC (Fig. 1c, middle panel). In each session, they performed four tasks resembling Experiment 1 after receiving TBS (Fig. 1c, middle and right panels). Slopes of curves were analyzed as in Experiment 1. We hypothesized that iTBS and cTBS upon the dLMC would improve and impair perception, respectively (Fig. 1d, right panel).

Results showed that, cTBS upon the dLMC inhibited both tone and consonant perception. Tone perception with noise masking was impaired by cTBS upon the left dLMC (compared with sham: $p_{fdr} = 0.029$, Cohen's d = -0.688, permutation test, Fig. 3c). Consonant perception was impaired by cTBS upon the left dLMC both when noise masking was absent (compared with sham: $p_{fdr} = 0.049$, Cohen's d = -0.474, permutation test, Fig. 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, Fig. 3c). cTBS upon the right dLMC 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, Fig. 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). We also tested the cTBS effects on psychometric slopes after removing "invalid slopes" instead of using slope replacement as we did in Fig. 3 (see Supplementary Information, Preprocessing of the slope). Results are shown in Fig. S2. In brief, cTBS effects were still significant after removing "invalid slopes" for both tone perception in noise and consonant perception in noise by left dLMC stimulation (Fig. 3c and Fig. S2g). Nevertheless, since inducing invalid curve fitting can also be regarded as a TMS disruption effect (for listen in noise conditions, TMS induced more invalid slopes compared with sham, Fig. S2e,f), we adopted the results after the preprocessing of slope replacement.

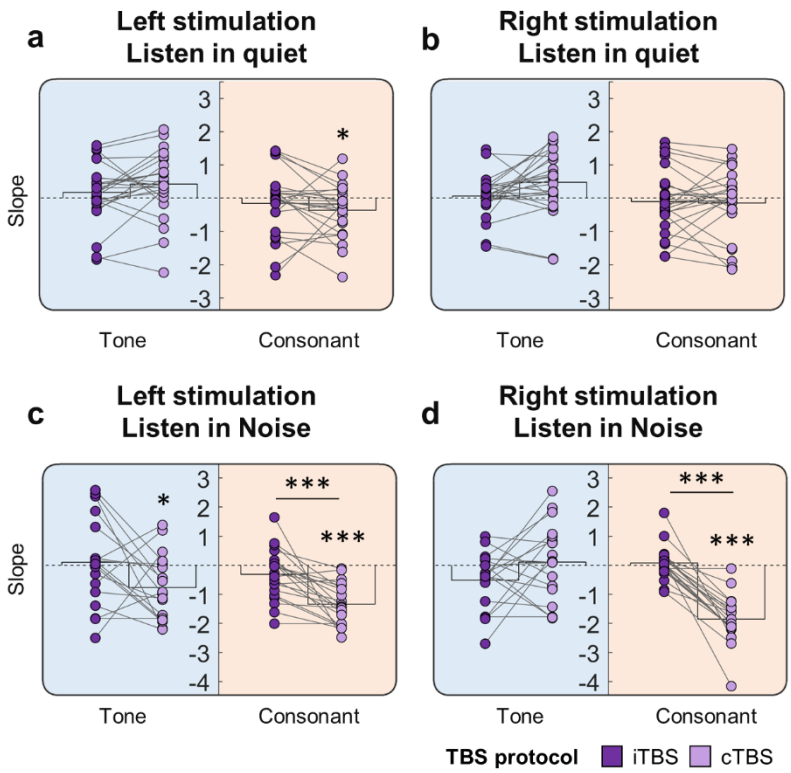

Figure 3. TBS effects upon the dLMC on slopes of tone and consonant identification curves in Experiment 2.

(**a**) cTBS upon the left dLMC impaired consonant perception in quiet.

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

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

(**d**) cTBS upon the right dLMC impaired consonant perception in noise.

See Fig. 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

(single-tailed) were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

The dLMC is causally involved in various psychological processes underlying perceptual decision-making

Next, we used the hierarchical Bayesian estimation of DDM (Hierarchical DDM)[26] to determine whether the dLMC is not only involved in the perception but also in the decision process, and if it does, what cognitive stages of the decision process it may participate in. DDM is a type of sequential sampling model that captures trial-wise variations in speed-accuracy tradeoff[21]. DDM assumes that perceptual evidence is noisily accumulated over time starting from a point of response bias (z), with an average drift rate (v), and triggers decision until one of the two boundaries (with a distance of threshold a) is reached. Hierarchical DDM allows simultaneously estimating subject-level parameters and group distributions, and provides posterior distributions of estimated parameters. Here, we modeled data in Experiment 2 and determined which parameters (a, v, or z) were affected by TBS by comparing 8 linear regression models recruiting none (baseline), one, two, or all (full) of the three parameters, respectively (see Methods, Drift-diffusion model and reaction time analysis).

Effects of iTBS were indistinct as models with parameters added did not consistently outperform the baseline models (Fig. S5 a–h), but the full models in cTBS conditions kept surpassing the baseline models (Fig. S5 i–p). We therefore focused on effects of cTBS. Since the full models in cTBS conditions had acceptable goodness-of-fit (RT distribution of the simulated data had a good similarity to that of the real data in the 95% credible criteria, see Methods, drift-diffusion model and reaction time analysis), we chose them as winning models and assumed that cTBS exerted effects on all three parameters. Hypotheses were tested by comparing posterior distributions of full model parameters in each condition with zero. As shown in Fig. 4 a-h (upper panels), compared with sham, cTBS significantly broadened the thresholds of decision boundary (a) for all the conditions (all posterior distributions $p_{fdr} < 0.05$, two-tailed) except the condition of tone perception in quiet with left dLMC stimulation. The drift

rates of evidence accumulation (v) were decreased by left dLMC stimulation in tone in quiet task (p_{fdr} = 0.014, two-tailed, Fig. 4a), and were increased by left dLMC stimulation in all the remaining conditions (all $p_{fdr} < 0.01$, two-tailed, Fig. 4b, e, f), but were not affected by right dLMC stimulation (Fig. 4c, d, g, h). In contrast, the starting points of evidence accumulation (response bias, z) were only affected in consonant in noise tasks by both left ($p_{fdr} < 0.001$, two-tailed, Fig. 4f) and right (p_{fdr} = 0.012, two-tailed, Fig. 4h) stimulation.

We then compared cTBS effects on DDM parameters and that on RTs. Using linear mixed-effects (lme) models, we found that cTBS significantly increased RTs (effects of cTBS vs. sham on the random intercept of the lme models, $ps_{fdr} < 0.05$) except for the tone in noise condition when the left dLMC was stimulated (Fig. 4 a–h Lower). This corresponds to the general increase in thresholds a as the decision boundary is closely related to the strategy of speed-accuracy tradeoff[27]. For RT analyses in iTBS conditions, see Fig. S6.

Meanwhile, the number of DDM parameters being altered by cTBS predicted the significance of the modulation on slopes of psychometric curves. At least two DDM parameters (a + v and/or z) were significantly modulated by cTBS in conditions with significant slope reduction (Fig. 4b, e, f, h, Upper, labeled in purple), whereas only one parameter (a or v) was modulated by cTBS in conditions without slope alteration (Fig. 4a, c, d, g, Upper, labeled in grey). Hence, clear consistency exists between the two independent analysis pipelines. This increases the confidence that we detected *authentic* TMS effects instead of artifacts induced by psychometric curve fitting and DDM modeling procedures (patterns generated by data processing would hardly be replicated in two completely different pipelines). Moreover, as there are relationships between the psychometric function and DDM parameters[28], we tried to combine results from the two pipelines and discuss what stages of perceptual decision mediated dLMC TMS effects on perception, or in other words, what stages the dLMC participate in (see Discussion for details).

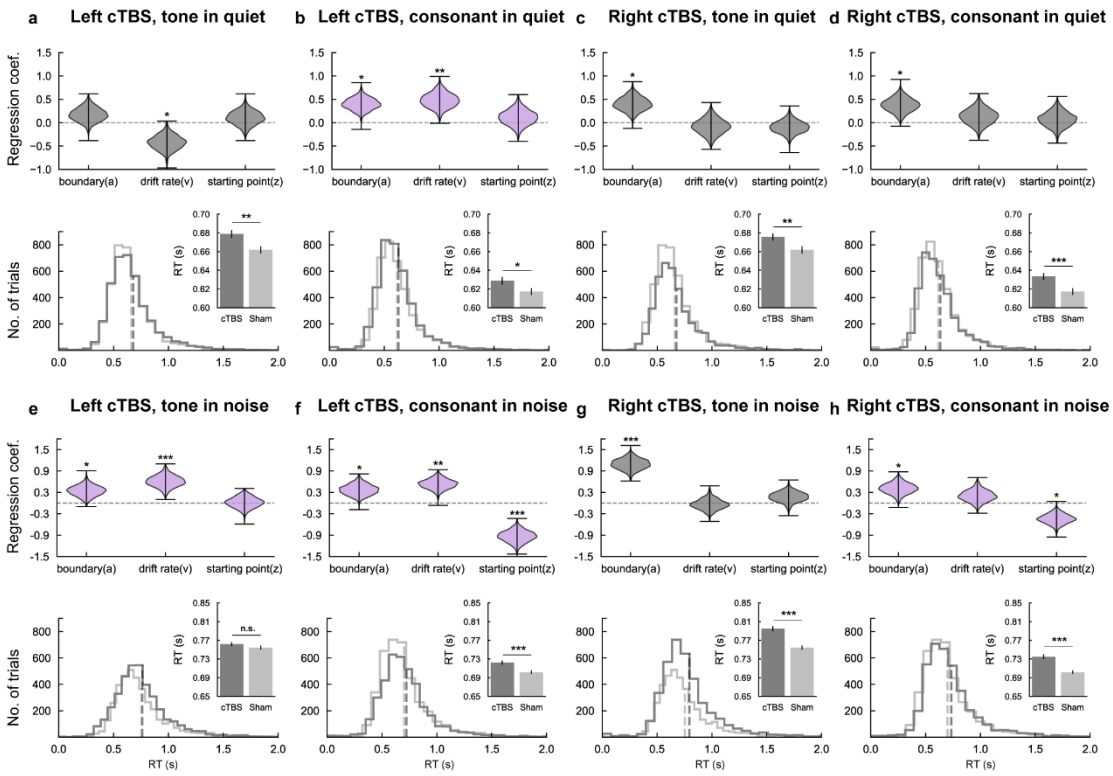

Figure 4. cTBS effects upon the dLMC on DDM parameters and RTs in Experiment 2.

(**a**) cTBS upon the left dLMC decreased DDM drift rates (upper panel, the same below), and lengthened RTs (lower panel, the same below) in tone perception in quiet.

(**b**) cTBS upon the left dLMC increased DDM thresholds of boundary and drift rates, and lengthened RTs in consonant perception in quiet.

(**c**) cTBS upon the right dLMC increased DDM thresholds of boundary, and lengthened RTs in tone perception in quiet.

(**d**) cTBS upon the right dLMC increased DDM thresholds of boundary, and lengthened RTs in consonant perception in quiet.

(**e**) cTBS upon the left dLMC increased DDM thresholds of boundary and drift rates in tone perception in noise.

(**f**) cTBS upon the left dLMC increased DDM thresholds of boundary, drift rates, and altered DDM starting points, and lengthened RTs in consonant perception in noise.

(**g**) cTBS upon the right dLMC increased DDM thresholds of boundary, and lengthened RTs in tone perception in noise.

(**h**) cTBS upon the right dLMC increased DDM thresholds of boundary and altered DDM starting points, and lengthened RTs in consonant perception in noise.

Upper panels: for all conditions, full DDM models assuming that cTBS affected all three parameters (*a*, *v*, and *z*) were selected and posterior distributions of parameters compared with sham were shown. Note that, conditions with slopes significantly impaired by cTBS (see Fig. 3) were shown in purple, whereas the other conditions were shown in grey.

Lower panels: for all conditions, single-trial RTs were compared between cTBS and sham using the lme with fixed slope (cTBS effect) and random intercept

(participant).

For all statistical tests, p values (two-tailed) were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

The dLMC engagement is left-biased and modulated by task difficulty

Lastly, we explored the hemispheric asymmetry of the dLMC involvement, and whether and how it is modulated by task difficulty. We integrated results from psychometric curve fittings and those from DDM modeling in Experiment 2. We focused on revealing systematic differences among indices (i.e., psychometric slope, and DDM parameters *a*, *v*, and *z*) between conditions with left/right dLMC stimulation and with/without noise-masking.

We found that the dLMC involvement is left-biased in terms of both tone and consonant perception. For slopes, tone perception was only hampered by cTBS upon the left dLMC and with noise-masking (Fig. 3c); consonant perception was inhibited by cTBS upon the left dLMC regardless of noise-masking (Fig. 3a, c), but was only affected by the right dLMC stimulation when noise was presented (Fig. 3d). For DDM parameters, cTBS affected drift rates (*v*) in all conditions with the left dLMC stimulated, but not in conditions with the right dLMC stimulated. In sum, the left dLMC weighs higher than its right counterpart, as cTBS upon the left dLMC induced greater changes in perceptual decision. However, both left and right dLMC cTBS lengthened RTs and widened DDM thresholds for decision boundary equivalently, suggesting that the right dLMC also makes contributions.

Masking noise increased the task difficulty and the engagement of bilateral dLMC. For slopes, only consonant perception in quiet was affected by the left dLMC stimulation (Fig. 3a), but perception in noise was more susceptible to cTBS (Fig. 3c, d). Note that, for slopes, stimulating the right dLMC affected consonant perception in noise (Fig. 3d), indicating a causal role of the right dLMC in adverse listening conditions.

Another factor in our study that affected the task difficulty was the ambiguity of stimuli (i.e., the step in a continuum), as stimuli with pitch contour (for tone) or VOT (for consonant) closer to categorical boundaries are more indistinguishable. We focused

on the DDM results for Experiment 2 and sought to determine whether interactions existed between stimulus ambiguity and cTBS effects on each parameter. We compared full models (cTBS affected *a*, *v*, and *z*) without interaction terms (baseline) and those with one, two, or all three (full) parameters interacting with the stimulus ambiguity (see the Supplementary Information, Interactions between cTBS effects and stimulus ambiguity for details). We found that across all cTBS conditions, the full models outperformed the baseline models and most of the remaining models. Simple main effect analyses for conditions with significant slope modulation showed that cTBS only affected the boundary (*a*) and the drift rate (*v*) when the stimuli were unambiguous (steps 1 and 5) or half-ambiguous (steps 2 and 4) (Table S2). In other words, the dLMC may not be engaged in perceptual decision when acoustic information is categorically ambiguous.

Discussion

Whether and how the LMC, especially the dLMC, is recruited in speech perception remains elusive. Here we provide causal evidence for shared representations of the dLMC in the perceptual decision of phonetic features with its roles in pitch regulation and voicing control during articulation. In Experiment 1, rTMS upon the dLMC, but not the TMC, inhibited both lexical tone and consonant perception. We applied cTBS upon the dLMC in Experiment 2 and confirmed the findings. For slopes of psychometric curves that represent perceptual sensitivity, left dLMC cTBS impaired tone perception in noise, and consonant perception in both quiet and noise; whereas right dLMC cTBS impaired consonant perception in noise. Thus, cTBS tended to exert greater effects when the left dLMC was stimulated, and when tasks were difficult (i.e., when stimuli were masked by noise). For DDM parameters that represent latent psychological processes in decision-making, cTBS widened thresholds for decision boundary (*a*) across all conditions, and increased drift rates (*v*) for conditions with the left dLMC stimulated. Also, cTBS changed starting points (*z*) in consonant perception in noise tasks, and rarely exerted effects when stimuli were completely ambiguous. Similar to the slope analyses, cTBS effects on DDM parameters were modulated by

stimulation hemisphere and task difficulty. The following text discusses the engagement of the dLMC in the perceptual decision of lexical pitch and voicing contrasts, their hemispheric functional specialization, as well as the effects of cognitive demands.

Lesion studies gave early causal evidence as patients reported impaired perception of voicing in consonant[29], lexical tone[30], and prosody[31] following damage to the frontal areas. More recently, "virtual lesion" studies by TMS have reported that the dLMC is causally involved in the discrimination of non-linguistic vocal pitch[32], singing voice[33], and paralinguistic prosody[34]. A recent study demonstrated that linguistic pitch (i.e., lexical tone) processing was suppressed after stimulating the left motor cortex[35]. Our study takes a step further and draws a causal link between the human dLMC (defined by BOLD activations in articulation tasks) and perceptual decision of lexical tone. Moreover, we provide substantial evidence that the dLMC 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 contrast[36].

In addition, we found that the engagement of the dLMC is left-biased but also recruits its right counterpart in perceptual decision of both linguistic pitch and voicing contrasts. Linguistic pitch (lexical tone) is a suprasegmental linguistic cue sharing identical articulatory acoustic essence with prosody and language functions with segmental phonemes[37]. The uniqueness of lexical tone has motivated research and fierce debate on its hemispheric asymmetry. Evidence has been provided for left-dominance[35,37], right-dominanc[38], and bilateral activations[39,40] in tone processing. The current study found that bilateral dLMC are involved in speed-accuracy tradeoff strategy selection (RTs and DDM threshold *a*) in tone identification. However, in terms of perceptual sensitivity, the dLMC engagement is left-lateralized. Notably, Tang et al. (2021)[35] discovered that tone perception in tonal and nontonal language speakers was separately hampered by the left and right motor cortex stimulation. Presumably, tonal language experience drives the motoric hub for tone perception from the right to the left hemisphere[41].

In terms of voicing contrasts (a type of consonant contrasts), consistent with the prevailing left-dominance models for phoneme perception[42,43], we found that motor simulation of voicing in consonants is biased to the left dLMC regarding its contributions to perceptual sensitivity. Indeed, the left speech motor cortex eases intra-hemispheric auditory sampling of phonemes[7]. Nevertheless, stimulating the right dLMC also inhibited consonant perception when noise was presented. Also, similar to tone perception, bilateral dLMC contributes equally to determining speed-accuracy tradeoff strategies. To our knowledge, this is the first study to report the participation of the right dorsal stream in the perception of fast-changing segmental linguistic units. Notably, the consonant results are inconsistent with Sammler et al. (2015)[34] as they found that stimulating the dLMC in neither side affected consonant perception. However, our tasks were more challenging due to extra noise masking, which might urge the dLMC engagement. Overall, previous studies have reported that the right motor cortex entrains speech stream[60] and is involved in auditory-speech coupling[61] resembling its left counterpart, and we provide new evidence for its engagement in the perception of lower-level phonemic cues.

For interactions between auditory and motor cortices in terms of functional asymmetry in speech perception, most studies have concentrated on functional distributions of the ventral pathways. 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)[52,53]. Hemispheric structural and functional differences of auditory cortices may account for this asymmetry (the *intrinsic* factors)[42,43,54]. In addition, individual differences in auditory asymmetry have been ascribed to the strength of auditory-frontal connectivity[55] (the *external* factors). Indeed, resembling the ventral pathways, the dorsal pathway for segmental linguistic units is left-dominant[4,35,37] whereas that for the suprasegmental paralinguistic prosody is biased to the right hemisphere[34,35]. Thus, it is speculated that the auditory asymmetry in speech perception may also be driven by the motor heterogeneity[35]. However, as the corpus callosum[56] and the fronto–temporo–parietal fiber tracts[57] provide the physiological basis for inter-

454 hemispheric cross-talk among bilateral auditory and motor cortices, this hypothesis is
455 still questionable. Nonetheless, we are not able to depict the functional interplay among
456 bilateral dorsal LMCs and the ventral pathways during speech perception (e.g.,[58])
457 merely based on our behavioral data. Future studies combining brain stimulation and
458 neuroimaging techniques are needed to fill these gaps.

459 Another phenomenon relevant to the hemispheric asymmetry is how the DDM drift
460 rates (v) were affected by cTBS. Only the left dLMC stimulation, but not the right,
461 affected drift rates. Note that, although the DDM assumes several independent
462 parameters in perceptual decision, cognitive and neural processes driving these
463 parameters are elusive. Previous studies have postulated that whereas the boundary
464 threshold is related to the ambiguity of sensory input[23,27], the drift rate is influenced by
465 both the stimuli saliency[23] and top-down prediction as it is increased by higher
466 dependence on contextual prior knowledg[44] from frontal region[27]. Our study found
467 bilateral effects on the boundary threshold and left-lateralized effects on the drift rate
468 by cTBS upon the dLMC. In addition, the number of DDM parameters modulated by
469 cTBS predicted the significance of slope modulation such that the slope was impaired
470 only when the DDM threshold boundary a was widened plus changes in the drift rate v
471 and/or starting point z. Methodologically, these correspondences validate the two
472 independent pipelines of data analyses. Psychologically, however, it is still difficult to
473 directly draw a link between perceptual sensitivity and latent decision processes. A
474 possible explanation is that, bilateral dLMC provide articulatory efference copy[45]
475 serving as categorical templates to match with sensory input. Categorical templates
476 increase the perceptual distance (i.e., sensitivity) between signals by non-linearly
477 transforming the psychological space[24]. Hence, when the dLMC was blocked,
478 perception became more uncertain. Faster drift rates after the left dLMC stimulation
479 may then reflect compensatory processes originating from a left-lateralized,
480 knowledge-based frontal system[46] that uses historical information to offset losses of
481 motor templates in perceptual decision. These compensatory processes were evident
482 especially when the motor template losses were sufficient to affect perceptual
483 sensitivity.

Interactions between cTBS effects and task difficulty point to a key debate on whether sensorimotor integration is essential[47] or modulatory[48] in speech perception. A related technical question is whether effects of TMS upon the motor cortex on behavioral performance reflects the essential role of the motor system in speech perception, or only changes in response bias as the hand motor area is perturbed[49,50]. Our study provides a more sophisticated perspective. First, our results support a modulatory view[5] as increased task difficulty (i.e., noise added) enhanced motor engagement (i.e., more cTBS effects on slopes). Moreover, when tasks are highly demanding (i.e., consonant in noise perception tasks where many or all participants performed unsatisfactorily; see Fig. S2), recruitment of bilateral dLMC became mandatory as participants relied heavily on subjective criteria of responses (response bias) after cTBS perturbed the left or right dLMC in consonant in noise tasks (reflected by changes in DDM starting point z). Furthermore, compensation from the dLMC disappears if perception becomes completely uncertain (i.e., no cTBS effects on a or v at ambiguous steps). This coincides with previous neuroimaging[16,17] and TMS[51] studies. Meanwhile, both the left and right dLMC is causally involved in the perceptual decision of lexical pitch and voicing, but may weigh differently. First, bilateral dLMC contribute equally to determining the threshold (a) for perceptual decision, presumably by providing articulatory motor templates to disambiguate sensory information. Second, the left dLMC seems to be a key node in perceptual decision vulnerable to virtual lesions, whereas the right dLMC is redundant and reluctant to perturbations until cognitive demands increase.

Limitations exist in the current study. Firstly, the location of the LMC is still debated. Although we did not find activation in ventral LMC (vLMC) for localization tasks, the vLMC has been suggested as the human homolog of the non-human primate LMC[62]. From the evolutionary perspective, a human-specific dLMC has emerged to handle vocal learning and control subserving language communications[63], through a hypothetical "duplication and migration" proces[64]. Thus, a human dLMC may implement fine-grained computations of speech vocalization[65]. Indeed, the ability to control vocal pitch during speech and singing is encoded by the dLMC in humans[13].

However, a recent study has also found that the vLMC is involved in both voicing production and perception[66]. Future studies are expected to directly compare causal contributions of the dLMC and vLMC to speech perception. Moreover, in daily life, perception of sentences with rhythmical, hierarchical, and contextual information may require more sophisticated functions of the motor cortex such as temporal[67], motor[68], and contextual prediction[69], as well as envelope tracking[70] in addition to perception of phonemic cues. Further studies are needed to investigate the functional relevance of the dLMC for speech perception in natural settings. Furthermore, our work is unable to demonstrate the neurobiological infrastructure for causal inferences for neurostimulation. After stimulation, short-term functional reorganization of neural networks may occur and complicate the causal interpretation[71].

To sum up, we applied TMS upon the dLMC of healthy participants which affected both the perceptual sensitivity and the decision processes of lexical tone and consonant perception. To our knowledge, this is the first study to provide causal evidence of the dLMC in speech perception and decision grounded in its articulatory functions (i.e., pitch contour and voicing control), paralleling the functions of lip (for bilabial features) and tongue (for dental features) motor areas. Moreover, bilateral dLMC are engaged with left-asymmetry, and their functions are modulated by cognitive demands such that their contributions are optimized at medium levels of difficulty. This study thus advances our knowledge of the effector-specific and modulatory role of the motor cortex in speech perception.

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. 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)[72], 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)[73]. 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.

Stimuli

Stimuli generation

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[36] and Chao's system of "tone-letters"[74]), corresponding to Mandarin syllables "堤 dī" (riverbank), "敌 dí" (enemy), "梯 tī" (ladder), and "题 tí" (title) (marked by Chinese Phonetic Alphabets "Pinyin"[22]), respectively (Audio S1). 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 noise-removed, re-concatenated and their pitch contours were normalized by Praat 6 (see Table S1 for acoustic properties of the syllables).

For Experiment 1, based on four natural syllables, 5 × 5 -step tone (tone1 to tone2)

570 and consonant ([t] to [t^h]) continuum matrices with individualized ranges of ambiguity
571 were generated by stepwise adjusting F0 and appending aspiration noise (Fig. 1b, left
572 panel). For Experiment 2, individualized 5-step tone ([ti55] to [ti35]) and consonant
573 ([ti55] to [t^hi55]) continua were generated in similar ways (Fig. 1c, left panel). To avoid
574 interactions between tone and consonant perception, no continuum matrix was used in
575 Experiment 2. See Supplementary information, Continua generation and Table S1 for
576 details.

577 **Stimuli presentation**

578 Stimuli were played by a Sennheiser in-ear headphones (IE 60) connected to a Steinberg
579 UR 242 amplifier. Intensity levels of stimuli were calibrated by Larson-Davis sound
580 level meter (Model 831, Depew, NY). In Experiment 1, syllables were presented by
581 Psychtoolbox-3 in Matlab 2016a at 78 dB sound pressure level (SPL), loud enough to
582 resist online TMS noise. In Experiment 2, syllables were presented at 65 dB SPL. This
583 was lower than that in Experiment 1 since no TMS noise affected stimuli presentation.
584 In conditions with noise-masking in Experiment 1, speech-spectrum noise (SSN)
585 started as a block started and was kept playing until the block ended. SSN was generated
586 by modulating white noise with the average spectrum of 5000 sentences spoken by 10
587 young female Mandarin speakers[75], and 4 kHz low-pass filtered. During the
588 presentation, the intensity of masking noise was adjusted to produce a signal-to-noise
589 ratio (SNR) of -4 dB, a medium level that was shown to induce maximal sensorimotor
590 integration in syllable decision[16,17]. In Experiment 2 we applied SSN identical to
591 Experiment 1 (except 5 kHz low-pass filtered) in conditions with noise-masking. The
592 intensity of SSN was adjusted to produce individualized SNRs. In each trial, SSN began
593 and ended simultaneously with the syllable, with additional 10 ms linear rise-decay
594 envelopes.

595 **MRI acquisition and processing**

596 For each participant in both experiments, a T1-weighted anatomical image for target
597 mapping was acquired by a 3-Tesla Siemens Magnetom Trio scanner with a 20-channel
598 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.

Experiment procedure

Experiment 1

Participants finished Experiment 1 on five separate days: day 1 for filling questionaries, assessing hearing and TMS thresholds, and preliminary experiments to estimate personal ranges of perceptual ambiguity of tone and consonant perception; day 2 for MRI acquisition; day 3 to 5 for rTMS experiments. rTMS experiments included 3 sessions: one sham session, one dLMC stimulation session, and one TMC stimulation session. The dLMC 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 (dLMC: [±40, -5, 50], TMC: [±59, -3, 36]; see Supplementary Information, 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 judgment (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 continuum 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.

Experiment 2

Participants finished Experiment 2 on seven separate days. Procedures for the first and second days were similar to Experiment 1. Day 3 to 7 were for TBS experiments, which included five sessions: sham, iTBS on the left dLMC, cTBS on the left dLMC, iTBS on the right dLMC, and cTBS on the right dLMC. 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

656 distributed) after participants' response or the automatic end of a trial. No break time
657 was left between adjacent mini-blocks. Identical to Experiment 1, participants
658 completed practice tasks before each TBS session.

659 **TMS**

660 **Experiment 1**

661 **Resting motor threshold estimation**

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

669 **TMS paradigm**

670 Participants were separated into two groups for left (N = 35, 20 females) and right (N
671 = 29, 16 females) hemisphere stimulation, respectively. Two groups of participants
672 were matched in age (left: mean = 22.43, SD = 3.03; right: mean = 21.90, SD = 2.67;
673 $t(62) = 0.73$, $p = 0.47$), RMT (left: mean = 61.09%, SD = 7.69; right: mean = 61.55%,
674 SD = 8.42; $t(62) = -0.23$, $p = 0.82$), and handedness (left: mean = 80.26, SD = 23.72;
675 right: mean = 80.44, SD = 20.60; $t(62) = -0.03$, $p = 0.98$).

676 During the experiment, triple-pulse rTMS was applied in synchronization with the
677 syllables. Biphasic magnetic pulses were generated by a standard figure-of-eight coil
678 from a Magstim Rapid2 stimulator. The intensity of stimulation was set to 90% of
679 participants' RMT. The frequency of stimulation was 10-Hz (100 ms between two
680 pulses). The first pulse was always delivered at the syllable onset, followed by two
681 subsequent pulses. Since a single TMS pulse induces aftereffects lasting for tens of
682 milliseconds[76], a triple-pulse sequence of stimulation can shadow an entire syllable.
683 Also, although it was suggested that high-frequency offline rTMS induces facilitatory

effects[77], 10-Hz online rTMS protocol has been shown to inhibit cognitive processin[78] and more specifically, down-regulate speech perception[79]. Presumably, while offline TMS alters performance by inducing short-term neural plasticity[80], online TMS affects cognition by modulating ongoing neural oscillations[81], 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 dLMC and TMC targets were mapped to the participants' T1-weighted images with Brainsight2, and were 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

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[82]. 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 contractio[83], 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[83]. 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 dLMC 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[83].

Quantification and statistical analysis

Experiment 1

Fitting psychometric function

Scores of identification tasks were binarily coded. 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. Slopes of psychometric functions served as the response variable. The lower the slopes, the weaker the participant's capacity to categorize phonemes.

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. Slopes were then preprocessed to reduce effects from invalid outliers (see Supplementary Information, Preprocessing of the slope and Fig. S1).

Stimulation effect analyses

Unexpectedly, we found competitions between tone and consonant perception such that slopes were affected by the ambiguity of the unattended dimension (see Supplementary Information, Competitions between consonant and tone perception). Thus, it is necessary to exclude such interference with rTMS effects. Therefore, for each block of tasks per session (dLMC 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 (Fig. S3 a).

rTMS effects were defined as differences between slopes in rTMS blocks and the corresponding sham blocks (rTMS - Sham). Also, the relative down-regulation of rTMS upon the dLMC on perception was estimated by comparing dLMC stimulation with TMC stimulation effects on slopes ((dLMC - Sham) - (TMC - Sham)). Details of statistical estimations are presented in Supplementary Information, Analysis of TMS effects and Fig. S1.

Experiment 2

Fitting psychometric function

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.

Drift-diffusion model and reaction time analysis

Single-trial data (binary decision and reaction times) were entered into DDM model [21] estimations. Here, we focused on three key parameters of the DDM model: the starting point or response bias (z), the average drift rate (v) of evidence accumulation, and threshold between decision boundaries (a). We conducted hierarchical Bayesian estimation of DDM using the HDDM package[26] based on Python 2.7.

First, we built 8 regression models recruiting none (baseline), one, two, or all (full) of the three parameters, respectively. We used dummy treatment coding with the intercept set on sham conditions. Each model drew 2000 posterior samples, and discarded the first 20 samples as burn-in. Models are shown below.

Baseline model: $a \sim 1$, $v \sim 1$, $z \sim 1$.

Model 1 (a only): $a \sim$ TBS, $v \sim 1$, $z \sim 1$.

Model 2 (v only): $a \sim 1$, $v \sim$ TBS, $z \sim 1$.

Model 3 (z only): $a \sim 1$, $v \sim 1$, $z \sim$ TBS.

Model 4 ($a + v$): $a \sim$ TBS, $v \sim$ TBS, $z \sim 1$.

Model 5 ($a + z$): $a \sim$ TBS, $v \sim 1$, $z \sim$ TBS.

Model 6 ($v + z$): $a \sim 1$, $v \sim$ TBS, $z \sim$ TBS.

Model 7 ($a + v + z$): $a \sim$ TBS, $v \sim$ TBS, $z \sim$ TBS.

We used the deviance information criterion (DIC, the lower the better)[84], a method widely used to assess and compare fit in hierarchical models, to select the winning model. A wining model should have a DIC value that is (1) lower than the baseline model and (2) consistently lower than most of the competing models in all conditions. For cTBS conditions, the model 7 (full model) that assumes cTBS effects on all three parameters was then chosen. For iTBS conditions, however, since no model met the criteria mentioned above, we did not conduct further analyses on the iTBS data (see Fig. S5 for comparisons of DIC values across models). Note that, a full model may

naturally perform better than the other models merely because it contains more parameters (i.e., over-fitting)[26]. Thus, we also simulated the data using the estimated parameters, and compared summary statistics between the real and the simulated data. Results showed that, the similarity fell in the 95% credible criteria for full models in all cTBS conditions.

Second, we tested whether cTBS significantly affected each of the three model parameters in each condition by comparing parameter posterior distributions to zero. Since we set the intercept of regression models on sham conditions, the probability that the posterior distribution is above or below zero (two-tailed) refers to the confidence (the p value) in rejecting the null hypothesis (cTBS exerted no effects). P values were then FDR-corrected for multiple comparisons across all conditions.

Third, we tested whether cTBS significantly altered RTs in each condition as a supplementary analysis for understanding how the dLMC influences strategies of speed-accuracy tradeoff. Since the data were hierarchical as single trials were nested in single participants, we used linear mixed-effects (lme) models to estimate differences in RTs between cTBS and sham conditions. Type of stimulation (cTBS or sham) was entered for the fixed slope, and participant was entered for the random intercept. We used the lme4 package in Rstudio to build the models (RTs ~ 1 + type of stimulation + (1|participant)), and used the lmerTest package to test the models. P values of the fixed slope (cTBS effects) were then FDR-corrected for multiple comparisons across all conditions.

Forth, to determine whether cTBS exerted equivalent effects on DDM parameters in trails with different levels of perceptual ambiguity (i.e., the acoustic position of the syllable in the consonant or tone continuum), we built DDM regression models with interaction items between the type of stimulation and the step in the consonant/tone continuum, and conducted simple main effect analyses if interaction existed. See the Supplementary Information, Interactions between cTBS effects and stimulus ambiguity and Fig. S7 for details.

Acknowledgements

The authors greatly appreciate the excellent work of the technical support staffs at the Institutional Center for Shared Technologies and Facilities of Institute of Psychology, Chinese Academy of Sciences. Also, 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.

This work was supported by the STI 2030—Major Projects 2021ZD0201500; the National Natural Science Foundation of China 31822024; the Strategic Priority Research Program of Chinese Academy of Sciences XDB32010300; and the Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences E2CX3625CX.

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. and 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	http://www.matlab.com	RRID: SCR_001622
Pyhon 2.7	http://www.python.org/	RRID:SCR_008394
R Project for Statistical Computing	http://www.r-project.org/	RRID:SCR_001905

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
BrainNet Viewer 1.7	https://www.nitrc.org/projects/bnv/	RRID: SCR_009446
Legacy STRAIGHT	https://github.com/ HidekiKawahara/legacy_STRAIGHT	RRID: SCR_021957
lme4	https://cran.r-project.org/web/packages/lme4/index.html	RRID:SCR_015654
lmerTest	http://CRAN.R-project.org/package=lmerTest	RRID:SCR_015656

Declaration of interests

The authors declare no competing interests.

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 available at GitHub (https://github.com/Baishenliang/TMS_dLMC_speech) before peer review. Custom code for the experiment, and visualization are available by request to the corresponding author.

References

1. Browman, C. P. & Goldstein, L. Articulatory gestures as phonological units. *Phonology* **6**, 201–251 (1989).
2. Brown, S., Ngan, E. & Liotti, M. A larynx area in the human motor cortex. *Cereb. Cortex* **18**, 837–845 (2008).
3. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
4. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
5. Hickok, G., Houde, J. & Rong, F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
6. Pulvermüller, F. & Fadiga, L. Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* **11**, 351–360 (2010).
7. Möttönen, R., Dutton, R. & Watkins, K. E. Auditory-motor processing of speech sounds. *Cereb. Cortex* **23**, 1190–1197 (2013).
8. Möttönen, R. & Watkins, K. E. Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* **29**, 9819–9825 (2009).
9. D' Ausilio, A. *et al.* The motor somatotopy of speech perception. *Curr. Biol.* **19**, 381–385 (2009).
10. D' Ausilio, A., Bufalari, I., Salmas, P. & Fadiga, L. The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* **48**, 882–887 (2011).
11. Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* **15**, 399–402 (2002).
12. D' Ausilio, A. *et al.* Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130418 (2014).
13. Dichter, B. K., Breshears, J. D., Leonard, M. K. & Chang, E. F. The control of vocal pitch in human laryngeal motor cortex. *Cell* **174**, 21–31 (2018).
14. Simonyan, K. The laryngeal motor cortex: its organization and connectivity. *Curr. Opin. Neurobiol.* **28**, 15–21 (2014).
15. Gick, B., Wilson, I. & Derrick, D. *Articulatory phonetics*. (Wiley-Blackwell, 2013).
16. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7126–7131 (2014).
17. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* **7**, 12241 (2016).
18. Du, Y. & Zatorre, R. J. Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 171222314 (2017).
19. Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T. & Adank, P. The effect of speech distortion on the excitability of articulatory motor cortex. *Neuroimage* **128**, 218–226 (2016).
20. Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the

speaking tongue: review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105 (2017).

21. Ratcliff, R. & Rouder, J. N. Modeling response times for decisions between two choices. *Psychol. Sci.* **9**, 347–356 (1998).
22. Huang, B. & Liao, X. *Xiandai Hanyu[Modern Chinese]*. (Higher Education Press, 2017).
23. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
24. Kronrod, Y., Coppess, E. & Feldman, N. H. A unified account of categorical effects in phonetic perception. *Psychon. Bull. Rev.* **23**, 1681–1712 (2016).
25. Suppa, A. *et al.* Ten years of theta burst stimulation in humans: established knowledge, unknowns and prospects. *Brain Stimul.* **9**, 323–335 (2016).
26. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. *Front. Neuroinform.* **7**, 1–10 (2013).
27. Mulder, M. J., van Maanen, L. & Forstmann, B. U. Perceptual decision neurosciences - a model-based review. *Neuroscience* **277**, 872–884 (2014).
28. Ratcliff, R. Measuring psychometric functions with the diffusion model. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 870–888 (2014).
29. Blumstein, S. E., Cooper, W. E., Zurif, E. B. & Caramazza, A. The perception and production of voice-onset time in aphasia. *Neuropsychologia* **15**, 371–383 (1977).
30. Gandour, J., Ponglorpisit, S. & Dardarananda, R. Tonal disturbances in Thai after brain damage. *J. Neurolinguistics* **7**, 133–145 (1992).
31. Witteman, J., Van Ijzendoorn, M. H., Van de Velde, D., Van Heuven, V. J. J. P. & Schiller, N. O. The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia* **49**, 3722–3738 (2011).
32. D' Ausilio, A., Bufalari, I., Salmas, P., Busan, P. & Fadiga, L. Vocal pitch discrimination in the motor system. *Brain Lang.* **118**, 9–14 (2011).
33. Lévêque, Y., Muggleton, N., Stewart, L. & Schön, D. Involvement of the larynx motor area in singing voice perception: a TMS study. *Front. Psychol.* **4**, 418 (2013).
34. Sammler, D., Grosbras, M.-H. H., Anwander, A., Bestelmeyer, P. E. G. & Belin, P. Dorsal and ventral pathways for prosody. *Curr. Biol.* **25**, 1–7 (2015).
35. Tang, D. L., Möttönen, R., Asaridou, S. S. & Watkins, K. E. Asymmetry of auditory-motor speech processing is determined by language experience. *J. Neurosci.* **41**, 1059–1067 (2021).
36. Ladefoged, P. & Johnson, K. *A Course in Phonetics*. (Michael Rosenberg, 2010).
37. Liang, B. & Du, Y. The functional neuroanatomy of lexical tone perception: an activation likelihood estimation meta-analysis. *Front. Neurosci.* **12**, 495 (2018).
38. Ge, J. *et al.* Cross-language differences in the brain network subserving intelligible speech. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2972–2977 (2015).
39. Si, X., Zhou, W. & Hong, B. Cooperative cortical network for categorical processing of Chinese lexical tone. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12303–12308 (2017).
40. Chien, P. J., Friederici, A. D., Hartwigsen, G. & Sammler, D. Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Hum. Brain Mapp.* **41**, 1842–1858 (2020).

41. Zatorre, R. J. & Gandour, J. T. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1087–1104 (2008).
42. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).
43. Zatorre, R. J., Belin, P. & Penhune, V. B. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46 (2002).
44. Davies, D. J., Teufel, C. & Fletcher, P. C. Anomalous perceptions and beliefs are associated with shifts toward different types of prior knowledge in perceptual inference. *Schizophr. Bull.* **44**, 1245–1253 (2018).
45. Tian, X. & Poeppel, D. Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Front. Hum. Neurosci.* **6**, 1–11 (2012).
46. Friederici, A. D. The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392 (2011).
47. Wilson, S. M. Speech perception when the motor system is compromised. *Trends Cogn. Sci.* **13**, 329–330 (2009).
48. Hickok, G., Holt, L. L. & Lotto, A. J. Response to Wilson: what does motor cortex contribute to speech perception? *Trends Cogn. Sci.* **13**, 330–331 (2009).
49. Smalle, E. H. M., Rogers, J. & Möttönen, R. Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cereb. Cortex* **25**, 3690–3698 (2015).
50. Venezia, J. H., Saberi, K., Chubb, C. & Hickok, G. Response bias modulates the speech motor system during syllable discrimination. *Front. Psychol.* **3**, 157 (2012).
51. Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T. & Adank, P. The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia* **94**, 13–22 (2017).
52. Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O. & Poeppel, D. Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Hum. Behav.* **3**, 393–405 (2019).
53. Albouy, P., Benjamin, L., Morillon, B. & Zatorre, R. J. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* **367**, 1043–1047 (2020).
54. Toga, A. W. & Thompson, P. M. Mapping brain asymmetry. *Nat. Rev. Neurosci.* **4**, 37–48 (2003).
55. Poeppel, D. & Assaneo, M. F. Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* **21**, 322–334 (2020).
56. Bloom, J. S. & Hynd, G. W. The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychol. Rev.* **15**, 59–71 (2005).
57. Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A. & Bolhuis, J. J. Language, mind and brain. *Nat. Hum. Behav.* **1**, 713–722 (2017).
58. Tamura, S. *et al.* Multi-modal imaging of the auditory-larynx motor network for voicing perception. *Neuroimage* 118981 (2022) doi:10.1016/j.neuroimage.2022.118981.
59. Park, H., Ince, R. A. A., Schyns, P. G., Thut, G. & Gross, J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**, 1649–1653 (2015).

60. Hincapié Casas, A. S. *et al.* Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. *Neuroimage* **244**, 118577 (2021).
61. Keitel, A., Ince, R. A. A., Gross, J. & Kayser, C. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage* **147**, 32–42 (2017).
62. Belyk, M. & Brown, S. The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77**, 177–193 (2017).
63. Jarvis, E. D. Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).
64. Mars, R. B., Eichert, N., Jbabdi, S., Verhagen, L. & Rushworth, M. F. S. Connectivity and the search for specializations in the language-capable brain. *Curr. Opin. Behav. Sci.* **21**, 19–26 (2018).
65. Eichert, N., Papp, D., Mars, R. B. & Watkins, K. E. Mapping human laryngeal motor cortex during vocalization. *Cereb. Cortex* **30**, 6254–6269 (2020).
66. Tamura, S. *et al.* Multi-modal imaging of the auditory-larynx motor network for voicing perception. *Neuroimage* **251**, 118981 (2022).
67. Rimmele, J. M., Morillon, B., Poeppel, D. & Arnal, L. H. Proactive sensing of periodic and aperiodic auditory patterns. *Trends Cogn. Sci.* **22**, 870–882 (2018).
68. Pickering, M. J. & Gambi, C. Predicting while comprehending language: a theory and review. *Psychol. Bull.* **144**, 1002–1044 (2018).
69. Brodbeck, C. *et al.* Parallel processing in speech perception with local and global representations of linguistic context. *Elife* **10**, e72056 (2022).
70. Ding, N. & Simon, J. Z. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* **107**, 78–89 (2012).
71. Bergmann, T. O. & Hartwigsen, G. Inferring causality from noninvasive brain stimulation in cognitive neuroscience. *J. Cogn. Neurosci.* **33**, 195–225 (2020).
72. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
73. Coffey, E. B. J., Herholz, S. C., Scala, S. & Zatorre, R. J. Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research. in *The Neurosciences and Music IV: Learning and Memory* (2011).
74. Chao, Y. R. A system of ‘tone-letters’. *Fangyan[Dialet]* **2**, 81–83 (1980).
75. Wu, X. *et al.* The effect of perceived spatial separation on informational masking of Chinese speech. *Hear. Res.* **199**, 1–10 (2005).
76. Sandrini, M., Umiltà, C. & Rusconi, E. The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neurosci. Biobehav. Rev.* **35**, 516–536 (2011).
77. Fitzgerald, P. B., Fountain, S. & Daskalakis, Z. J. A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clin. Neurophysiol.* **117**, 2584–2596 (2006).
78. Beynel, L. *et al.* Effects of online repetitive transcranial magnetic stimulation (rTMS) on cognitive processing: a meta-analysis and recommendations for future studies. *Neurosci. Biobehav. Rev.* **107**, 47–58 (2019).
79. Hartwigsen, G. *et al.* Dissociating parieto-frontal networks for phonological and semantic word decisions: a condition-and-perturb TMS study. *Cereb. Cortex* **26**, 2590–2601

(2016).
80. Thickbroom, G. W. Transcranial magnetic stimulation and synaptic plasticity: experimental framework and human models. *Exp. Brain Res.* **180**, 583–593 (2007).
81. Thut, G. *et al.* Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Curr. Biol.* **21**, 1176–1185 (2011).
82. Mas-herrero, E., Dagher, A. & Zatorre, R. J. Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. *Nat. Hum. Behav.* **2**, 27–32 (2018).
83. Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P. & Rothwell, J. C. Theta burst stimulation of the human motor cortex. *Neuron* **45**, 201–206 (2005).
84. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. Bayesian measures of model complexity and fit David. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64**, 583–639 (2002).

Supplementary Information

Continua generation

Experiment 1

Continua were synthesized by Matlab R2016a and legacy-STRAIGHT. Each continuum included multiple steps of syllables, 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] [1]), 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.

5×5-step individualized continuum matrices were generated to fit participants' ranges of perceptual ambiguity (steps in a continuum where participants were unable to 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 **Table S1** 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. Afterward, to ensure that participants understood the tasks, they finished a discrimination task to determine if the heard pair of syllables shared an identical tone/consonant category. The discrimination task was continued until participants discriminated all pairs correctly.

Experiment 2

The 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.

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.

Behavioral data analysis

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 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 **Figure S1** for a graphical illustration of this procedure).

Analysis of TMS effects

TMS effects were defined as differences between slopes in TMS blocks and the corresponding sham blocks, and were statistically tested using single-tailed one-sample permutation test in Matlab 2016a (if TMS - Sham < 0 or not). Compared with t-test, permutation test does not require the normality assumption of sample distribution, as distributions of slopes in some conditions were skewed as estimated the by Kolmogorov-Smirnov test of normality. To perform a permutation test, for each condition in TMS blocks, we randomly selected participants and permuted their slopes in the TMS block and sham block, contrasted TMS 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 TMS effect among values in the permutation distribution: $p = (\text{ranking} + 1) / (10^5 + 1)$. In addition, for Experiment 1, the relative down-regulation

of rTMS upon the dLMC on categorical perception was estimated by comparing dLMC stimulation with TMC stimulation effects on slopes ((dLMC - Sham) - (TMC - Sham)) using single-tailed Wilcoxon signed-rank test in Matlab 2016a. For Experiment 2, the relative down-regulation of cTBS upon the dLMC compared with iTBS was assessed in similar ways ((cTBS - Sham) - (iTBS - Sham)). Here, we used Wilcoxon signed-rank test instead of permutation test because the different TMS conditions (dLMC and TMS stimulation in Experiment 1, and iTBS and cTBS stimulation in Experiment 2) shared the same sham condition, and it was unable to simultaneously permute the sham condition with two TMS conditions. To correct for multiple comparisons, FDR correction was used with p threshold = 0.05. In Experiment 1, a group of comparisons refers to one group of trials (ambiguous, half-ambiguous, and unambiguous) in each group of participants (stimulating left or right hemisphere). In Experiment 2, a group of comparisons refers to all conditions the group of participants underwent.

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 (steps 2 and 4), and unambiguous (steps 1 and 5) conditions according to the unattended dimension of a matrix (**Figure S3A**). 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 (steps 2, 3, 5, and 6), and unambiguous (steps 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 (steps 2 and 4), and unambiguous (steps 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 (**Figure S1**). 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 the Kolmogorov-Smirnov test of normality showed that slopes in the unambiguous condition were not normally distributed, the 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 $\eta^2 = 0.103$; Pretest 1: $F_{2,30} = 5.001$, $p = 0.013$, partial $\eta^2 = 0.250$; 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 [2], and such competition was amplified when the unattended feature was half-ambiguous as participants had biased but uncertain perception.

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., dLMC 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 dLMC, respectively. The experiment was in a block design and was divided into 8 blocks, including 4 TMC and 4 dLMC 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

Preprocessing of anatomical and functional images were implemented by fMRIprep, an fMRI preprocessing pipeline that incorporates multiple state-of-art software tools [3].

T1-weighted (T1w) anatomical images were corrected for intensity non-uniformity and skull-stripped with ANTs 2.3.3 [4] (RRID:SCR_004757), and used as T1w-reference. Spatial normalization to standard space was performed through nonlinear registration with ANTs 2.3.3. The template used for spatial normalization was FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model [5] (RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym).

To process functional images, fMRIprep was used to correct for susceptibility distortions. Co-registrations of the BOLD to anatomical references were performed by ANTs 2.3.3. Head-motion parameters were estimated by FSL 5.0.9 [6]. Slice-timing correction was performed using AFNI 20160207 [7] (RRID:SCR_005927). The BOLD time-series were resampled and realigned to the MNI152NLin6Asym standard space. AFNI 20220424 was applied for spatial smoothing (Gaussian filter FWHM = 6.0 mm) and scaling the BOLD time-series to a mean of 100 to derive percent signal change.

Generalized linear model (GLM) analyses were conducted at the individual level by AFNI 20220424. 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 (dLMC: "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 thresholded by multiple comparison corrections (3dttest++ ClustSim program in AFNI 20220424, uncorrected $p < 0.001$, 10000 Monte Carlo simulations, smallest cluster size for "AH" – "D" = 239 voxels, "D" – "AH" = 164 voxels), and were masked by ROIs of bilateral precentral gyrus in automated anatomical labeling (AAL) templates [8] (RRID:SCR_003550).

MNI coordinates of bilateral homologous dLMC 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. As for results, the spatial patterns of activations were similar to previous studies [9,10], except that we did not find group-level significant activations in the ventral LMC. Whether the phonation area is located in the dorsal or the ventral or both of the LMC is still debated, but studies and models have emphasized the role of the dorsal LMC in human voice and pitch regulations [11] (See **Discussion** for details). Thus, the current study only focuses on the dorsal LMC. The chosen MNI coordinates were: [±40, -5, 50] for dLMC and [±59, -3, 36] for TMC (**Figure S4**). 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 [12]. Thus, transmission effects of stimulation can be ignored.

Drift-diffusion model and reaction time analysis

Interactions between cTBS effects and stimulus ambiguity

To study ***whether*** cTBS effects on DDM parameters were affected by stimulus ambiguity (i.e., the step in a continuum), we built 8 regression models with various interaction items based on the full model that assumes cTBS affected all three parameters (the winning model, see **Methods, Drift-diffusion model and reaction time analysis**). These models recruit none (baseline), one, two, or all (full) of the three parameters, respectively. Note that the "baseline" model here is equal to the full model without interaction items. We used dummy treatment coding with the intercept set on sham conditions. Each model drew 2000 posterior samples, and discarded the first 20 samples as burn-in. Models are shown below (AMB: stimulus ambiguity).

Baseline model:

a ~ cTBS, v ~ cTBS, z ~ cTBS.

Model 1 (*a* only):

a ~ cTBS + cTBS:AMB, v ~ cTBS, z ~ cTBS.

Model 2 (*v* only):

a ~ cTBS, v ~ cTBS + cTBS:AMB, z ~ cTBS.

Model 3 (*z* only):

a ~ cTBS, v ~ cTBS, z ~ cTBS + cTBS:AMB.

Model 4 (*a* + *v*):

a ~ cTBS + cTBS:AMB, v ~ cTBS + cTBS:AMB, z ~ cTBS.

Model 5 (*a* + *z*):

a ~ cTBS + cTBS:AMB, v ~ cTBS, z ~ cTBS + cTBS:AMB.

Model 6 (*v* + *z*):

a ~ cTBS, v ~ cTBS + cTBS:AMB, z ~ cTBS + cTBS:AMB.

Model 7 (*a* + *v* + *z*):

a ~ cTBS + cTBS:AMB, v ~ cTBS + cTBS:AMB, z ~ cTBS + cTBS:AMB.

We used the deviance information criterion (DIC) to select the winning model. A winning model should have a DIC value that is (1) lower than the baseline model and

(2) consistently lower than most of the competing models in all conditions. **Figure S7** for comparisons of DIC values across models. Results showed that the model 7 (full model) where cTBS affects all three parameters and all effects interact with stimulus ambiguity had the lowest DIC values in most conditions, and hence it was chosen as the wining model. Moreover, the same as the cTBS effect analyses, we simulated the data based on the estimated parameters, and compared summary statistics between the real and the simulated data. The similarity fell in the 95% credible criteria for full models in all conditions. Therefore, cTBS effects on all three DDM parameters were affected by stimulus ambiguity.

We then did simple main effect analyses to study ***how*** cTBS effects on three DDM parameters were modulated by stimulus ambiguity. For each condition, trials were separated into three groups according to the step in the tone/consonant continuum (steps 1 and 5: unambiguous; steps 2 and 4: half ambiguous; step 3: ambiguous). For each group of trials, an independent full model that assumes cTBS affected all three parameters was built and posterior distributions of parameters (compared with sham) were tested. Model and parameter estimations were the same as main effect analyses of TBS effects (see **Materials and methods**). Results were shown in **Table S2**. Note that for the boundary threshold (a) and the drift rate (v), significant cTBS effects on parameter distributions were only shown when stimuli were not completely ambiguous (i.e., half or un-ambiguous). Thus, it is possible that when sensory information provides absolutely no valid features for perceptual decision, the dLMC loses its contributions.

Supplementary figures and tables

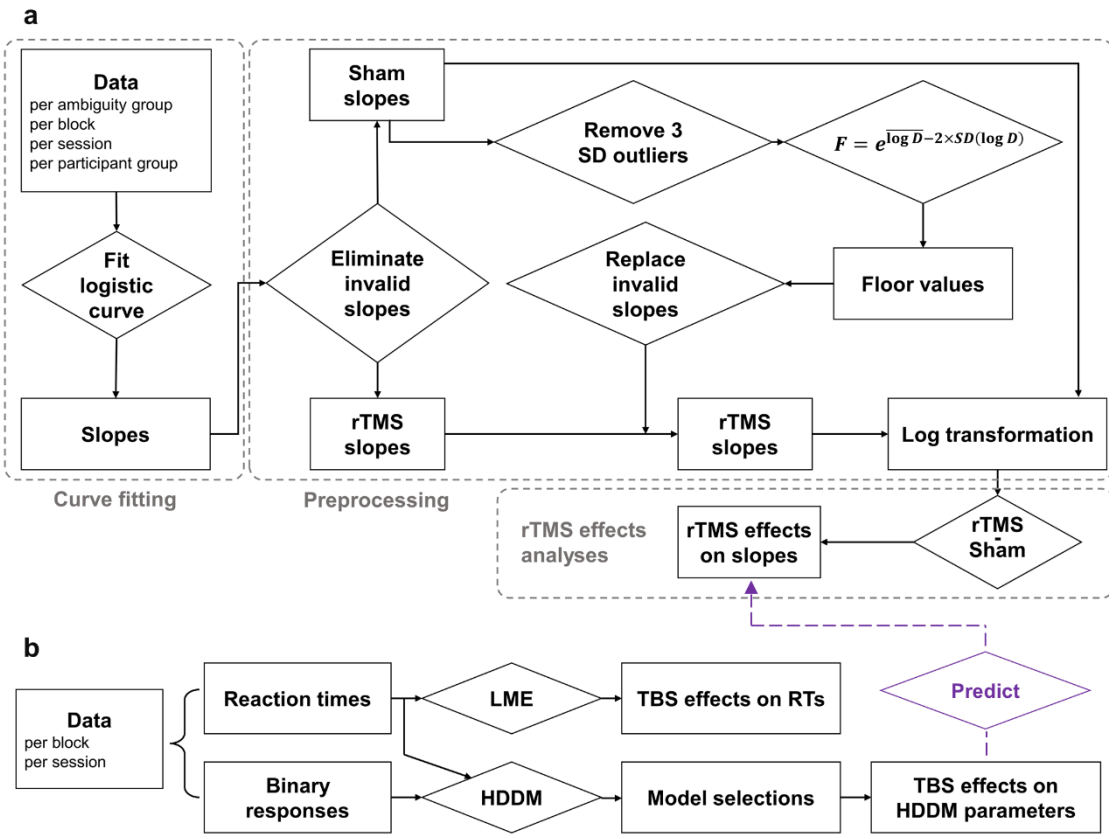

Figure S1. Data analysis pipelines.

(a) Perceptual sensitivity analysis includes curve fitting, preprocessing of slopes, and rTMS effects analyses (for Experiment 1; those for Experiment 2 were similar).

(b) Latent psychological processes analysis includes modeling, model selections, and post-hoc analyses of TBS effects on model parameters. TBS effects on reaction times were also estimated using linear mixed model estimation.

Besides, we also seek to explain changes in psychometric curve slopes by DDM parameters through comparing results from the two analysis pipelines (purple text).

LME: linear mixed model estimation; HDDM: hierarchical Bayesian estimation of the drift-diffusion model.

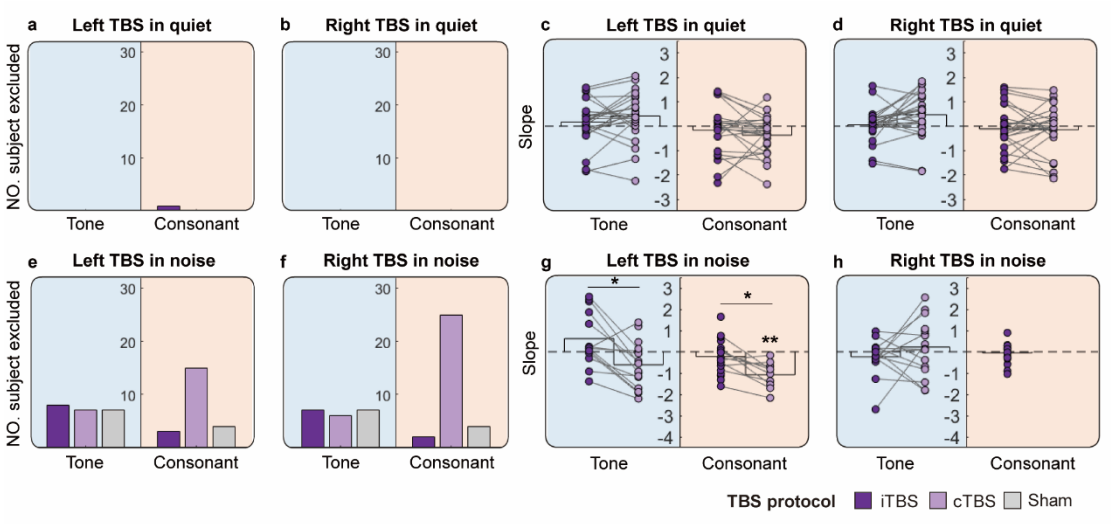

Figure S2. Elimination of invalid slopes and results without slope replacement in Experiment 2.

(a – b) and (e – f) Number of participants excluded in each condition for whom the psychometric slopes were marked as "invalid" (See **Materials and methods**, and **Figure S1** for details). Note that more invalid results were found in conditions with noise presented, especially for consonant perception in noise (**e – f**).

(c – d) and (g – h) Results of psychological curve slopes in Experiment 2 before slope replacement (See **Materials and methods**, and **Figure S1** for details). The same results but with slope replacement are shown in **Figure 3**.

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 (single-tailed) were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

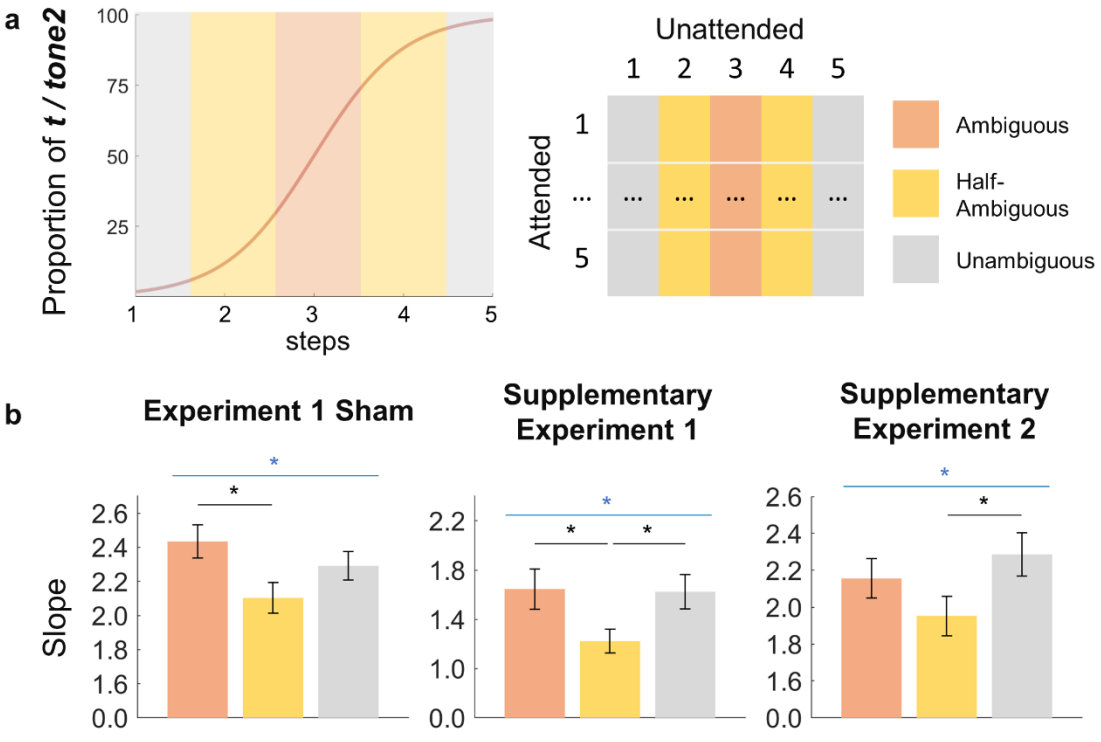

Figure S3. Competitions between tone and consonant perception.

(**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 judgments at unambiguous steps, biased but uncertain judgements at half-ambiguous steps, and guessing judgments 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 $p < 0.05$.

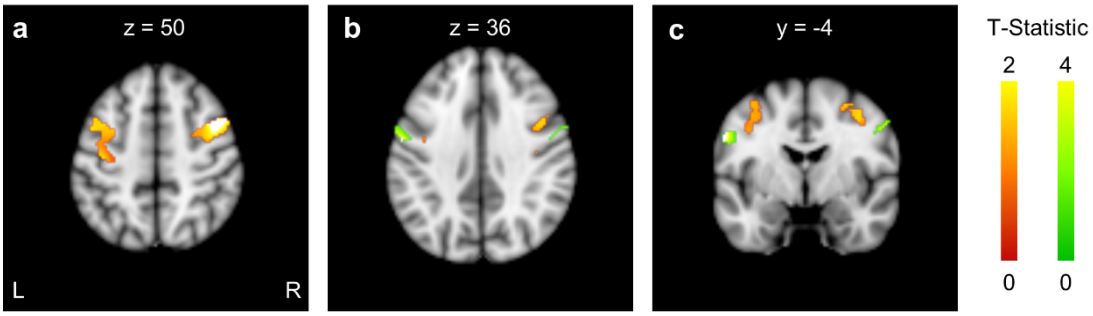

Figure S4：Group-level activation in the functional localization experiment.

Activation areas were masked by AAL ROIs of the bilateral precentral gyrus. The axial views showed the position of dLMC (**a and b**, orange) and TMC (**b**, green). **(c)** Dorsal-ventral functional dissociation between larynx (orange) and tongue (green) motor representation was shown in the coronal view of activation. L, left hemisphere; R, right hemisphere. z and y represent MNI coordinates at the inferior-to-superior and anterior-to-posterior extents, respectively. Visualization was performed by Mango [13].

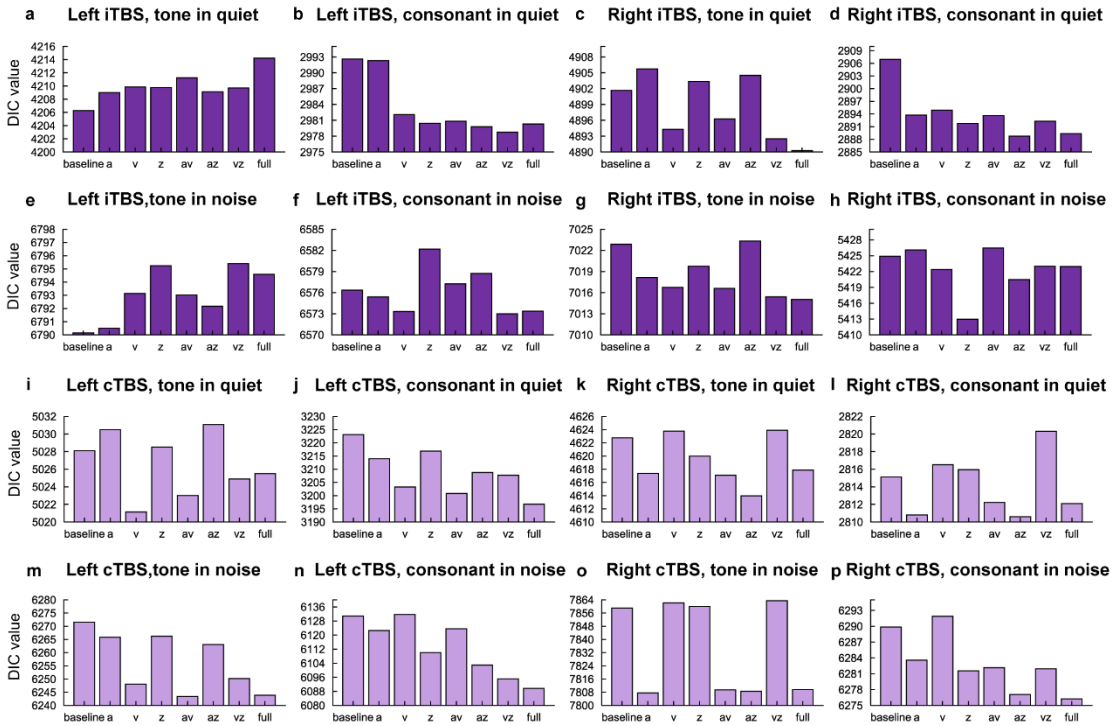

Figure S5：HDDM deviance information criterion (DIC) values of eight different models in various conditions.

(a – h) Model comparisons in iTBS conditions.

(i – p) Model comparisons in cTBS conditions.

DDM parameters: a: threshold for decision boundaries; v: drift rate for evidence accumulation; z: the starting point of evidence accumulation or response bias.

Baseline: baseline regression model (null model) that assumes no TBS effects; a/v/z: models assuming only one of the three DDM parameters is affected by TBS; av/az/vz: models assuming two of the three DDM parameters are affected by TBS; full(avz): the model assuming all three DDM parameters are affected by TBS.

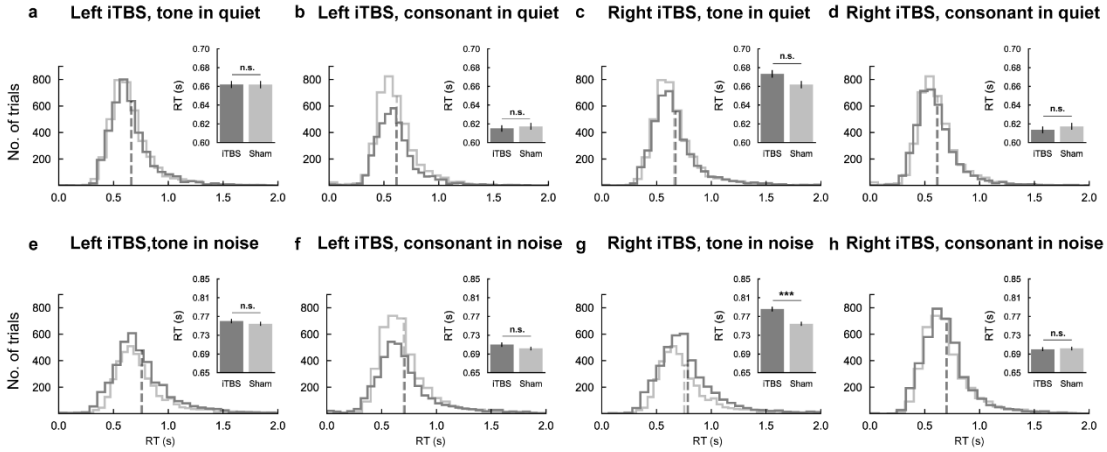

Figure S6：iTBS effects upon the dLMC on RTs in Experiment 2.

Compared with sham, iTBS upon the right dLMC increased RTs in the tone perception task in noise (**g**), but did not affect RTs in the other conditions (**a–f, h**).
For all statistical test, p values (two-tailed) were adjusted by false discovery rate (FDR) correction (threshold = 0.05). * $p_{fdr} < 0.05$; ** $p_{fdr} < 0.01$; *** $p_{fdr} < 0.001$.

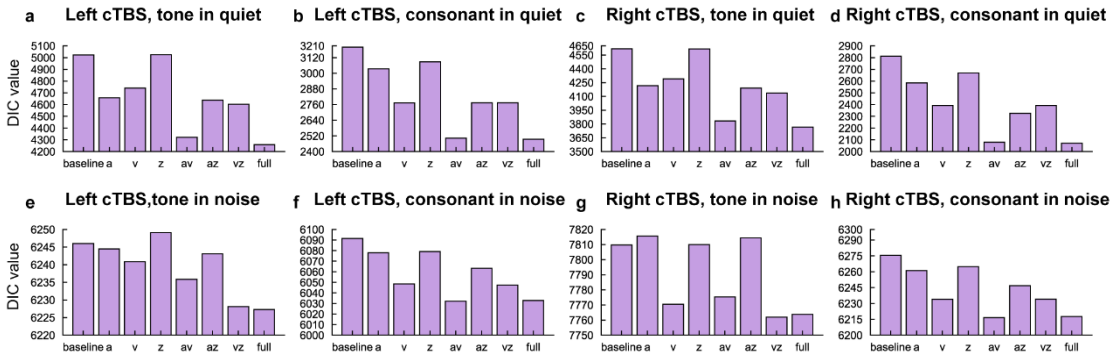

Figure S7：HDDM deviance information criterion (DIC) values of eight different models with interaction items in various conditions.

DDM parameters demonstrated are equivalent to **Figure S5**.

Note that the "baseline" model here is equal to the "full" model that assumes cTBS affects all three parameters but without interactions with stimulus ambiguity; a/v/z: models that assume cTBS affects all three parameters and only one of the effects interacts with stimulus ambiguity; av/az/vz: models that assume cTBS affects all three parameters and two of the effects interact with stimulus ambiguity; full(avz): the model assuming cTBS affects all three parameters and all effects interact with stimulus ambiguity.

Table S1. 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

Table S2. Interactions between effects from cTBS and ambiguity of continua step on DDM parameters.

Conditions where cTBS exerted significant effects on slopes of psychometric functions are analyzed only. Simple main effect analyses separately in trials with unambiguous, half-ambiguous, and ambiguous tone/VOT are performed if the interaction is significant. +/- signs and numbers represent directions and posterior probabilities of regression coefficients of interactions, or cTBS effects in simple main effect analyses compared with sham (+: higher than sham, -: lower than sham). Probabilities are FDR corrected. Numbers in bold show significant effects (posterior probability p < 0.05, two-tailed). Posterior probability: ’ $p < 0.1$ (marginally significant), * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Boundary (a)				Drift rate (v)				Starting point (z)			
	Inter.	**UAMB**	**HAMB**	**AMB**	**Inter.**	**UAMB**	**HAMB**	**AMB**	**Inter.**	**UAMB**	**HAMB**	**AMB**
Left cTBS Consonant in quiet	+ **0.000*****	+ 0.056’	+ **0.048***	0.576	+ **0.000*****	0.888	+ **0.000*****	0.325	+ **0.016***	0.244	- 0.444	0.6
Left cTBS Tone in noise	+ **0.006****	0.243	0.308	0.576	+ **0.000*****	+ **0.000*****	+ **0.000*****	- 0.860	**-** **0.044***	- 0.244	**-0.012***	+ **0.036***
Left cTBS Consonant in noise	+ **0.000*****	0.328	+ **0.048***	- 0.932	+ **0.000*****	+ **0.040***	+ 0.056’	0.325	- 0.096’			
Right cTBS Consonant in noise	+ **0.000*****	+ 0.08’	+ 0.056’	0.608	+ **0.000*****	0.237	+ 0.056’	- 0.325	- 0.186			

Supplementary references

1. Huang, B. & Liao, X. *Xiandai Hanyu[Modern Chinese]*. (Higher Education Press, 2017).
2. Luber, B. & Lisanby, S. H. Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *Neuroimage* **85**, 961–970 (2014).
3. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
4. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
5. Evans, A. C., Janke, A. L., Collins, D. L. & Baillet, S. Brain templates and atlases. *Neuroimage* **62**, 911–922 (2012).
6. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
7. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10**, 171–178 (1997).
8. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).
9. Brown, S., Ngan, E. & Liotti, M. A larynx area in the human motor cortex. *Cereb. Cortex* **18**, 837–845 (2008).
10. Dichter, B. K., Breshears, J. D., Leonard, M. K. & Chang, E. F. The control of vocal pitch in human laryngeal motor cortex. *Cell* **174**, 21–31 (2018).
11. Belyk, M. & Brown, S. The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77**, 177–193 (2017).
12. Bolgnini, N. & Ro, T. Transcranial magnetic stimulation: disrupting neural activity to alter and assess brain function. *J. Neurosci.* **30**, 9647–9650 (2010).
13. Lancaster, J. L. *et al.* Automated regional behavioral analysis for human brain images. *Front. Neuroinform.* **6**, 1–12 (2012).