Check for updates

The motor system's [modest] contribution to speech perception

Ryan C. Stokes 1 · Jonathan H. Venezia 1 · Gregory Hickok 1

Published online: 3 April 2019

© The Psychonomic Society, Inc. 2019

Abstract

Recent evidence suggests that the motor system may have a facilitatory role in speech perception during noisy listening conditions. Studies clearly show an association between activity in auditory and motor speech systems, but also hint at a causal role for the motor system in noisy speech perception. However, in the most compelling "causal" studies performance was only measured at a single signal-to-noise ratio (SNR). If listening conditions must be noisy to invoke causal motor involvement, then effects will be contingent on the SNR at which they are tested. We used articulatory suppression to disrupt motor-speech areas while measuring phonemic identification across a range of SNRs. As controls, we also measured phoneme identification during passive listening, mandible gesturing, and foot-tapping conditions. Two-parameter (threshold, slope) psychometric functions were fit to the data in each condition. Our findings indicate: (1) no effect of experimental task on psychometric function slopes; (2) a small effect of articulatory suppression, in particular, on psychometric function thresholds. The size of the latter effect was 1 dB (~5% correct) on average, suggesting, at best, a minor modulatory role of the speech motor system in perception.

Keywords Speech · Motor theory of speech perception · Motor cortex · Articulatory suppression

Introduction

Speech perception requires the listener to rapidly decode complex acoustic patterns. Traditionally, auditory regions in the superior temporal lobe have been implicated in this task (Binder et al., 2000; Klatt, 1980; Kuhl & Miller, 1971). On the other hand, speech production, or the precise coordination of articulatory movements, has been classically linked to brain regions in the left inferior frontal lobe (Price, 2000). These systems may not act independently of one another; some scientists propose that the auditory system supports speech production (Guenther, Hampson, & Johnson, 1998; Hickok, Houde, & Rong, 2011; Houde & Jordan, 1998), while others assert that the motor system is required for optimal speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Wilson, 2009; Wu, Chen, Wu, & Li, 2014). The first claim will not be

addressed in this paper. The second claim, derived from the motor theory of speech perception, is driven by evidence demonstrating motor system involvement during speech recognition tasks (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Meister et al., 2007; Watkins, Strafella, & Paus, 2003; Wu et al., 2014).

Motor theory of speech perception

Liberman et al. proposed that the objects of speech perception are not acoustic patterns, but rather vocal tract gestures produced by the speaker (Liberman et al., 1967; Liberman & Mattingly, 1985). The motor theory of speech perception (henceforth "motor theory") claims that the listener maps variable acoustic speech inputs to invariant gestural (i.e. motor) commands to recover the identity of incoming speech sounds. This theory is attractive, in part because it provides solutions to the perceptual problems posed by coarticulation, which causes neighboring speech sounds to blend into one another in the acoustic speech stream. Several iterations of the motor theory have been proposed, including a strong version and two weak versions.

The strong version of the motor theory claims that the objects of speech perception are vocal tract gestures, which must engage the motor system of the perceiver in order for speech perception to occur (Liberman et al., 1967). While evidence does show motor system activation during

Department of Cognitive Sciences Social and Behavioral Sciences Gateway, University of California - Irvine, Irvine, CA 92697-5100, USA



Ryan C. Stokes stokesr@uci.edu

perception of speech (Watkins et al., 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004), this alone does not provide evidence that motor activation is necessary or even helpful in deciphering the speech signal. When there is damage to motor speech systems, such as in cases of Broca's aphasia, patients often exhibit speech production deficits, yet are still capable of accurately perceiving speech sounds (Hillis, 2007; Mohr et al., 1978). Several other conditions, including congenital anarthria, cerebral palsy, bilateral inferior frontal lesions, and left hemisphere anesthesia, destroy or prevent the normal functioning of the speech motor system without compromising speech perception (Bishop, Brown, & Robson, 1990; Craighero, Metta, Sandini, & Fadiga, 2007; Hickok, 2010; Hickok et al., 2008; Lenneberg, 1962). Additionally, prelinguistic infants and some animal species are able to discriminate speech sound categories while lacking the motor repertoire required to reproduce what they hear (Kuhl & Miller, 1971; Werker & Yeung, 2005). Thus, in normal listening conditions, there is little evidence to support the notion that the motor system plays a crucial role in speech perception.

The first of the weak versions was a revision advanced by Liberman himself, which proposes that *intended* gestures are the objects of speech perception (Liberman & Mattingly, 1985). However, the representational nature of an "intended gesture" is unclear. Given the mounting evidence that the targets of motor gestures are in fact auditory in nature (Guenther et al., 1998), it would not be logically inconsistent under this weaker version of the motor theory to propose that acoustic representations are the intended objects of speech perception.

More recently, another even weaker version of the motor theory has been proposed, based on converging experimental evidence from studies involving speech perception in difficult listening situations (Wu et al., 2014). This version claims the motor system influences speech perception, but it is not sufficient or essential for perception, suggesting instead that the motor system plays an ancillary role in which it is primarily engaged during noisy or degraded listening scenarios (Wilson, 2009; Wu et al., 2014). Additional support comes from a computational model of speech recognition in which auditory and motor theories were found to be indistinguishable during ideal listening conditions but not during degraded conditions (Barnaud, Bessière, Diard, & Schwartz, 2018; Laurent, Barnaud, Schwartz, Bessière, & Diard, 2017). Henceforth, we will focus on this weaker version of the motor theory which seems to be at the heart of most recent scientific literature on the subject.

Functional magnetic resonance imaging (fMRI) (Watkins et al., 2003; Wilson et al., 2004) and transcranial magnetic stimulation (TMS) studies (Fadiga et al., 2002; Meister et al., 2007; Möttönen & Watkins, 2012; Panouillères, Boyles, Chesters, Watkins, & Möttönen, 2018) have been used to asses the role of motor activation during speech perception. In an fMRI experiment, Wilson et al. presented

speech sounds during passive listening and found activation in a portion of the premotor cortex belonging to the mouth motor system (Wilson et al., 2004). This indicates some level of functional connection between auditory and motor speech systems, but provides no direct evidence that motor system activation facilitates speech perception. On the contrary, TMS studies have provided such causal evidence, reporting a decrease in subjects' ability to perform syllable perception tasks when repetitive TMS is applied to the premotor or motor areas (D'Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011; Meister et al., 2007; Möttönen & Watkins, 2009). For instance, by using TMS to disrupt the premotor cortex, Meister et al. (2007) demonstrated that participants performed worse in a phonetic discrimination task compared to the same task without disruption.

While these TMS studies represent the best evidence for the weaker version of the motor theory, the effect size in these studies tends to be small and effects disappear when speech is presented in quiet listening conditions (D'Ausilio et al., 2014). Therefore, these results contradict the strong version of the motor theory and provide consistent, but not compelling, evidence for the weaker version. Additionally, response bias and/or task effects may be contributing to the already modest effects observed (Venezia, Saberi, Chubb, & Hickok, 2012). Specifically, participants may use motor resources in order to attend to phonemic information in the discrimination tasks that are otherwise not engaged in naturalistic speech perception (Hickok, 2012b). The methods used in Meister et al. present an additional complication - namely, speech sounds were presented in noise at a constant SNR (Meister et al., 2007). This is a problem because, as the SNR changes, performance in two different conditions (e.g., TMS vs. no-TMS) may change at different rates. We already know that performance approaches normal as the SNR is increased to an arbitrarily high level (i.e., quiet) (D'Ausilio et al., 2014). To best understand how performance is affected at intermediate SNRs, including identification of the SNRs for which the motor system is maximally engaged, performance should be measured across a wide range of SNRs rather than at a single value.

Therefore, the goal of the present study was to measure the effects of vocal tract motor suppression on phoneme identification across a wide range of SNRs. Participants performed a minimal-pair phoneme identification task in noise. The SNR was varied adaptively and the parameters of a logistic psychometric function were estimated from the data in order to characterize the full range of performance across SNRs.

Psychometric functions

The psychometric function provides a quantitative description of the relationship between a stimulus parameter, in this case the SNR of speech in noise, and performance on a behavioral task (here, minimal-pair phoneme identification). As the SNR



increases, we expect to see an increase in performance (percent correct). This relationship can be modeled using a logistic function with four parameters:

$$y = \gamma + \frac{(1 \text{-} \gamma \text{-} \lambda)}{\left(1 + exp(-\beta(x \text{-} \alpha))\right)}$$

The threshold (α) indicates the center of the psychometric function with a slope proportional to β . The guess rate (γ) indicates the lower horizontal asymptote of the function. In this case, since we are using a two-alternative forced choice task (75% threshold), we can assume the guess rate is 50%. The lapse rate (λ), or the expected percentage of misses as SNR reaches its highest levels, represents the upper horizontal asymptote (ceiling or near ceiling performance).

Example

Psychometric functions provide a more complete view of performance differences across conditions. Consider two hypothetical psychometric functions (Fig. 1) with an experimental condition (red), and a control condition (blue). If we set the functions to have the same threshold, but a different slope (Fig. 1a), then an experiment that tests a single SNR at X_1 will report a positive effect (greater percent correct in experimental vs. control), an experiment that tests a single SNR at X_3 will report a negative effect, and an experiment that tests a single SNR at X₂ will report a null effect. While accurate, these single-SNR findings are incomplete and potentially misleading considering the entire range of measurable values. If we now set our psychometric functions to have the same slope, but different thresholds, we run into a different problem. Single-SNR experiments at X₄ (large threshold shift, matched but shallow slopes) and X₅ (small threshold shift, matched but steep slopes) will report the same effect size, even though the underlying threshold shift is different. These considerations are nontrivial, but remain overlooked in the literature.

We explored the extent of motor system involvement in speech perception by estimating psychometric functions in four conditions of a phoneme identification task: (1) during articulatory suppression by means of subvocal repetition; (2) while opening and closing the mandible; (3) while tapping the foot; and (4) during passive listening (i.e., no secondary task). Articulatory suppression involves subvocal repetition of a word, which behaviorally disrupts the motor system during a speech perception task, comparable to an application of TMS to motor speech brain regions. Participants were asked to repeat the word the to occupy the speech motor system. The word the was selected because it is content free, high frequency, and monosyllabic. Similar words or speech sounds are used in the articulatory suppression literature to minimize non-motor processing (Baddeley, Lewis, & Vallar, 1984; Hanley & Bakopoulou, 2003; Liu, Squires, & Liu, 2016; Saeki & Saito, 2004). In the mandible condition, participants were asked to open and close their jaw without any vocalization, actual or imagined, to control for potential interference via auditory imagery in the articulatory suppression condition (Sams, Möttönen, & Sihvonen, 2005). The foot tapping condition served as a control for non-speech motor effects on speech perception (i.e., measured the demands of performing a secondary motor task in general). Passive listening provided a baseline for comparison. Participants performed minimalpair phoneme identification based on distinctions in both voicing and place of articulation. Place of articulation and voicing distinction tasks differ in their range of perceptibility (Alwan, Jiang, & Chen, 2011; Cole, Jakimik, & Cooper, 1978) allowing us to work with a wider range of SNRs. We used an adaptive staircase procedure to guide SNR selection, and estimated both the threshold and slope of the psychometric function from the data in each experimental condition.

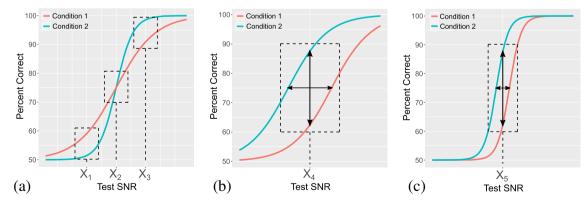


Fig. 1 (a) Three example experiments that test the effect of condition (percent correct, vertical difference) at three different SNR levels (X_1, X_2, X_3) when the threshold of a psychometric function is the same between conditions but the slope is different. X_1 reports negative results, X_2 null results, and X_3 positive results. (b) Sample experiment

 X_4 where a large threshold shift is observed between conditions with a fixed shallow slope. (c) Sample experiment X_5 , in which conditions have a fixed steep slope, will yield the same effect size as experiment X_4 despite a much smaller threshold shift



Psychometric functions allowed us to make comparisons along the entire range of SNR values and more completely delineate the effects of motoric suppression on speech perception in noise. This analysis is crucial to determine whether and how the motor system contributes to speech perception.

Methods

As a preregistered report, the methods and analysis were planned and submitted before the experiment was carried out. The experiment code, collected data, and analysis can be found online: https://osf.io/tqhr5/

Participants

Twenty-four participants were recruited from the University of California at Irvine student population through the Human Subjects Lab Pool; data from 4 participants were excluded from analysis (see Methods - Procedure for excluding data). Data from the remaining 20 participants (age 20.5 ± 4.1 years; 17 females) were submitted to further analysis. All participants received course credit for their participation. Participants were required to be at least 18 years old with normal, or corrected to normal vision and normal hearing by self-report.

Our target sample size was 20 participants. To ensure a low probability of a type II error, we calculated power using the methods described in "Statistical Power for the Two-Factor Repeated Measures ANOVA" (Potvin & Schutz, 2000) in conjunction with the power analysis software program G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to obtain a conservative estimate. The calculations assumed .90 power at the .05 alpha error probability using an effect size and correlation matrix based on pilot data.

Stimuli

Four minimal phonemic pairs consisting of monosyllabic, consonant-vowel English words were selected (Buy/Pie, Die/Tie, Buy/Die, and Pie/Tie) as the speech stimuli; two were distinguishable on the basis of place of articulation distinction (bilabial vs. alveolar: Buy/Die and Pie/Tie) and two were distinguishable on the basis of voicing (voiced vs. unvoiced: Buy/Pie and Die/Tie). Speech stimuli were created by digitally recording each word from a female speaker at a sampling rate of 44,100 Hz. A single CV syllable was then chosen as a model to generate synthesized versions of all four syllables. The formant structure and glottal source of the selected sound were extracted using a linear predictive coding algorithm in Praat (Boersma & Weenink, 2016). Formant trajectories were manually altered and used to filter the model source. Specifically, the slopes of the formant transitions of F2 and

F3 were adjusted to values that best approximated the desired place of articulation distinction. Then, the "burst" portions of the original waveforms for voiced (Buy/Die) and unvoiced (Pie/Tie) stimuli were excised and appended to the beginning of the synthetic tokens at zero crossings in the waveform. To create unvoiced tokens (Pie/Tie), aspiration noise from the original unvoiced tokens was averaged and appended between the burst and formant transitions of the corresponding synthetic tokens at zero crossings in the waveforms. The resulting synthetic speech sounds were then matched for overall duration using the pitch synchronous overlap-add (PSOLA) method in Praat and normalized to equal root-mean-square amplitude. Noise was generated by filtering Gaussian noise to match the long term average spectrum of the female speaker. The level of the speech-shaped noise was held constant at a comfortable listening level (~70 dB SPL), while the level of the speech signals was varied to produce different SNRs. Pairs of images corresponding to the meaning of each tested minimal pair were presented on a computer screen to cue participant responses.

Procedure

The experiment was conducted in the MATLAB environment using Psychtoolbox version 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Auditory stimuli were presented over headphones and visual images of the corresponding word pairs were presented on the left and right sides of a laptop screen, respectively (target location determined at random). Responses were indicated with the "Z" and "M" keys of the laptop keyboard corresponding to the left- and right-hand images, respectively. The experiment was split into 16 blocks, with each block comprising one of four conditions (articulatory suppression, mandible movement, foot tapping, and passive listening) paired with one of the four minimal pairs (pseudorandom order). For the articulatory suppression blocks, participants were instructed to repeat the word the subvocally but with overt movement of the vocal tract articulators - every second for the duration of the block. For the mandible movement condition, participants were instructed to open and close their jaw every second. For the foot tapping condition, participants were instructed to tap their foot every second. Finally, for the passive listening blocks, participants were instructed to listen to the stimuli without subvocal rehearsal and without motor movement.

Each block contained four 2-down, 1-up staircase runs targeting a 70.7% correct performance level. A single speech sound (429 ms) embedded in the center of a segment of speech shaped noise (1000 ms) was presented at the start of each trial (Fig. 2). The noise was presented at a constant and comfortable level. Speech sounds were presented at an initial level such that participants could achieve ceiling performance. The SNR was subsequently decreased with a step size of 6



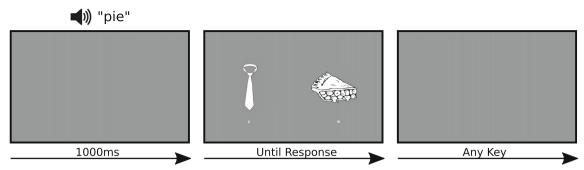


Fig. 2. A sample trial sequence. A speech sound in noise is presented for 1000 ms followed by a forced choice task

dB until the first reversal, whereby the the step size was changed to 4 dB. After the next reversal, the step size was set to 2 dB for the remainder of the procedure. Each staircase run consisted of 50 trials. Thus, a total 200 trials were collected for each experimental condition. Simulations suggest this number of trials is sufficient to estimate the slope of the psychometric function from adaptive data (Leek, Hanna, & Marshall, 1992). The adaptive track was restarted for each 50 trial run to ensure adequate sampling of the upper "elbow" of the psychometric function. After the audio stimulus, both words were visually presented on the screen until the participant made a selection, which was followed by a blank screen. Participants were given practice at the beginning of the experiment until they felt comfortable with the instructions. At the end of each block, participants were given the opportunity to take a break.

Procedure for excluding data. If any of the 4 staircase procedures in each condition did not converge to the threshold (by visual inspection of staircase trajectories), the data from that staircase were excluded from further analysis. If more than one staircase did not converge within a block, that participant's entire dataset was excluded from analysis.

Results

The staircase procedure ensures that near-threshold regions of the psychometric function are sampled the most often (Fig. 3). High SNR values will have few samples and high accuracy, while values close to threshold will have more samples, represented in Fig. 3 as larger dots, and have greater weight during estimation of the psychometric function. A psychometric function was fit for each participant across all combinations of minimal pair (Pie/Buy, Die/Tie, Buy/Die, and Pie/Tie) and task (articulatory suppression, mandible movement, foot tapping, and passive listening) for a total of 16 unique conditions. Palamedes toolbox (Prins & Kingdon, 2009) was used to perform maximum likelihood estimation of the best fitting psychometric function with two free parameters: threshold and slope. The guess rate was held constant at 0.5 reflecting the probability of a random selection being correct. The lapse

rate was held constant at 0.02 to mitigate bias in the estimate of the psychometric function slope (Klein, 2001). The parameter estimates for each subject in each condition were averaged across participants to generate a set of group-average psychometric functions (Fig. 4).

A two-way repeated measures ANOVA was used to compare the effect of the minimal pair condition and the task conditions on the slope parameter. *Task* (articulatory suppression, mandible movement, foot tapping, and passive listening) was the first factor and *minimal pair* (Pie/Buy, Die/Tie, Buy/Die, and Pie/Tie) was the second factor. The main effect of *minimal pair* was significant (F(3.0, 57.0) = 5.8, p = 0.002), but neither the main effect of *task* nor the *task x minimal pair* interaction were significant.

Since the slope parameter varied significantly across the minimal pairs but not tasks, we chose to fix the slope parameter across all conditions within each minimal pair. The fixed values were set to the average slope across all participants and conditions for a given minimal pair, thus a new set of psychometric function fits were generated with a fixed slope parameter for each of the four minimal pairs (Table 1). A repeated measures ANOVA with factors task and minimal pair was then carried out on the threshold estimates from the newly fitted, fixed-slope psychometric functions. There was a statistically significant task x minimal pair interaction (F(4.5,86.1) = 2.8, p = 0.026, Greenhouse/Geisser corrected). Simple main effect analysis of task within each minimal pair revealed a significant difference only for the Die/Tie minimal pair (F(3.0, 57.0) = 5.6. p = 0.002). Post-hoc comparisons performed between all pairs of tasks (Bonferroni correction) showed that participants performed significantly worse in the foot tapping and mandible movement conditions compared to the articulatory suppression condition (p = 0.040and p = 0.027, respectively). Participants also performed significantly worse in the mandible movement condition compared with the passive listening condition (p = 0.015).

Post-hoc analyses involving direct comparisons between the passive listening condition and each of the dual-task experimental conditions were also carried out. These analyses were intended to increase sensitivity to the relatively small effects of the dual-task conditions relative to the passive



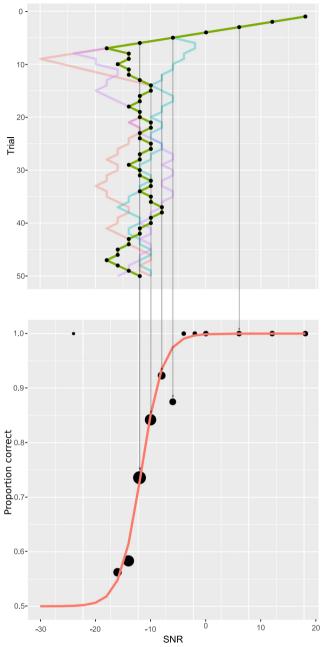


Fig. 3. For each condition and subject, a psychometric function was estimated by using the data from up to four adaptive staircase runs. This method allows for larger amounts of data to be sampled around the most informative parts of the psychometric function. A greater number of trials at a given SNR on the top plot is represented as larger circles on the bottom plot.

listening control. To this end, four separate repeated measures ANOVAs were run with *threshold* as the dependent variable, task as the first factor, where the two levels were one of the four pairwise combinations of passive listening with a single dual-task experimental condition, and *minimal pair* as the second factor. When task consisted of passive listening vs. articulatory suppression (effect of speech motor suppression), the main effects of task (F(1.0, 19.0) = 4.5, p = 0.047) and

minimal pair (F(3.0, 57.0) = 195.2, p < 0.001) were significant, but the two-way interaction was not significant. Performance was poorer for articulatory suppression compared to passive listening (average effect = 1.0 dB). When task consisted of passive listening vs. mandible movement (effect of non-speech vocal motor interference), the two-way interaction (F(3.0, 57.0) = 3.7, p = 0.017) and the main effect of minimal pair (F(1.9, 35.5) = 200.6, p < 0.001) were significant, but the main effect of task was not significant. The simple main effects of task were investigated within each minimal pair, which revealed a significant difference between mandible movement and passive listening only for the Die/Tie minimal pair (F(1.0, 19.0) = 12.0, p = 0.003). Performance was significantly poorer for the mandible movement compared to passive listening for Die/Tie (average effect = 2.0 dB). When task consisted of passive listening vs. foot tapping (non-vocal motor interference), the main effect of minimal pair was significant (F(3.0, 57.0) = 154.8, p < 0.001), but the main effect of task and the two-way interaction were not significant.

Discussion

The central aim of this study was to test what we have termed the "weaker" version of the motor theory of speech perception, which asserts that the speech motor system is engaged in perception primarily as a compensatory mechanism during noisy or otherwise difficult listening conditions. Thus, the study aimed to characterize the effect of speech-motor suppression on minimal-pair phoneme identification in speechshaped noise. This was accomplished via a dual-task behavioral manipulation in which participants were asked to subvocally-but-overtly articulate (i.e., moving the vocal tract articulators without phonating) during phoneme identification (articulatory suppression). In a second dual-task condition, participants made overt, up-and-down mandible movements during phoneme identification (mandible movement). In the mandible movement condition, participants were instructed not to generate any overt or covert speech, thus ensuring that effects on performance were not due to speech-specific motor interference or associated auditory-speech imagery. In a third dual-task condition, participants performed rhythmic tapping of the foot during phoneme identification (foot tapping). This condition was intended to capture the effect of general (nonvocal-tract) motor interference on performance. Finally, participants performed phoneme identification in a baseline condition with no concurrent task (passive listening). Performance was measured across four minimal phoneme pairs encompassing two place distinctions (Buy/Die; Pie/ Tie) and two voicing distinctions (Pie/Buy; Die/Tie). Crucially, the SNR was varied adaptively across a range of values, allowing for estimation of complete psychometric



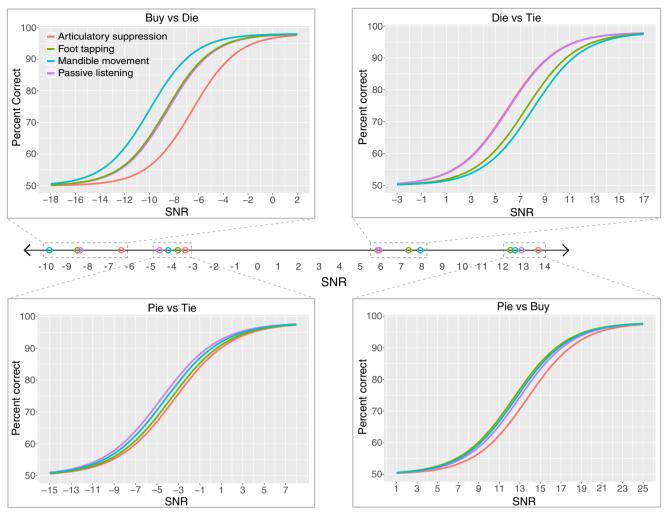


Fig. 4. Psychometric functions for the fixed slope model of all 16 conditions. The threshold of each condition is plotted on the center SNR number line, surrounded by a boundary box which contains all

tasks for each minimal pair. The full psychometric function of each task is drawn within the corresponding subplot

functions for each combination of minimal pair and task condition. Two-parameter psychometric functions were characterized by their threshold (75% correct performance) and slope (sensitivity of performance to changes in SNR) allowing for a comprehensive depiction of task effects across minimal pairs.

The results can be summarized as follows: (1) the psychometric function slope did not vary systematically as a function of task, but did vary significantly across minimal pairs. Voicing (Buy/Pie, Die/Tie) distinctions were found to be more difficult to distinguish compared to place of articulation conditions (Buy/Die, Pie/Tie). The brief, low-amplitude segment of aspiration noise used to convey voice onset time in the voicing conditions was likely masked by the speech-spectrum background noise even at high SNR levels. This made voicing distinctions more difficult overall, which allowed us to test over wide range of SNRs across place and voicing conditions; (2) the psychometric function threshold varied as a function of task and

minimal pair, but the effects of task tended to be small and somewhat variable across listeners and minimal pairs; (3) the only task condition for which the psychometric function threshold was consistently higher (i.e., worse performance) than passive listening was articulatory suppression; and (4) the average effect size of articulatory suppression relative to passive listening was ~1 dB across minimal pairs. From this, we can conclude that speech-motor interference produced a specific, but quite modest, decrement in speech perception performance. This decrement was apparent as a rightward shift of the entire psychometric function with no concomitant change in slope, which, given the range of slopes observed across minimal pairs, would produce, at best, a roughly 13% decrease in performance (i.e., percent correct) at a given SNR on the quasilinear portion of the psychometric function (Fig. 5). Such a pattern is consistent with a genuine loss of perceptual sensitivity as opposed to effects of (in)attention, increased internal noise



Table 1 Parameters of psychometric functions (mean +/- 95% confidence interval). Within-subject 95% confidence intervals are reported (Morey, 2008)

Pair	Task	Variable slope		Fixed slope	
		Threshold	Slope	Threshold	Slope
Pie/Buy	Articulatory suppression Passive listening	13.74 ± 0.88 12.89 ± 0.77	0.41 ± 0.11 0.37 ± 0.08	13.73 ± 0.84 12.92 ± 0.80	0.39
	Foot tapping	12.37 ± 0.99	0.43 ± 0.12	12.38 ± 0.99	
	Mandible movement	12.36 ± 1.27	0.34 ± 0.07	12.61 ± 1.05	
Die/Tie	Articulatory suppression Passive listening	5.48 ± 1.03 5.99 ± 0.95	$0.49 \pm 0.13 \\ 0.55 \pm 0.11$	5.95 ± 0.92 6.02 ± 0.95	0.49
	Foot tapping	7.15 ± 0.71	0.54 ± 0.09	7.44 ± 0.84	
	Mandible movement	7.76 ± 0.75	0.40 ± 0.13	8.01 ± 0.95	
Buy/Die	Articulatory suppression Passive listening	-7.18 ± 1.38 -9.19 ± 0.91	$\begin{array}{c} 0.41 \pm 0.13 \\ 0.61 \pm 0.16 \end{array}$	-6.50 ± 1.71 -8.49 ± 1.21	0.55
	Foot tapping	-8.91 ± 1.74	0.55 ± 0.14	-8.62 ± 1.92	
	Mandible movement	-10.23 ± 1.69	0.62 ± 0.13	-9.99 ± 2.01	
Pie/Tie	Articulatory suppression Passive listening	-3.92 ± 1.17 -4.90 ± 0.76	$\begin{array}{c} 0.37 \pm 0.11 \\ 0.31 \pm 0.11 \end{array}$	-3.39 ± 1.44 -4.64 ± 0.88	0.37
	Foot tapping	-4.45 ± 1.17	0.38 ± 0.09	-3.75 ± 1.49	
	Mandible movement	-4.60 ± 1.38	0.43 ± 0.10	-4.22 ± 1.44	

and/or fluctuations in the decision criterion, or any other task element that interacts with SNR, all of which are expected to produce a shift in the slope of the psychometric function under most circumstances (Buss, Hall III, & Grose, 2009; Kontsevich & Tyler, 1999; Morgan, Dillenburger, Raphael, & Solomon, 2012).

Convergence with existing data

Of the existing data concerning the effect of motor suppression on speech perception, the most compelling evidence was reported by Meister et al. (2007) who showed that repetitive

(disruptive) TMS applied to vocal-tract-premotor cortex produced an average reduction in phoneme-discrimination-innoise performance of 8.3% correct (average baseline performance = 78.9% correct). This effect falls squarely within the range of effects observed presently for articulatory suppression (Fig. 5) and is thus consistent with a 1-2 dB shift in the psychometric function. We have argued previously (e.g. Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011; Venezia et al., 2012) that Meister et al.'s data could plausibly be explained by changes in "task-level" aspects of performance such as response selection, phonological working memory, or their interaction. As noted above, the present data

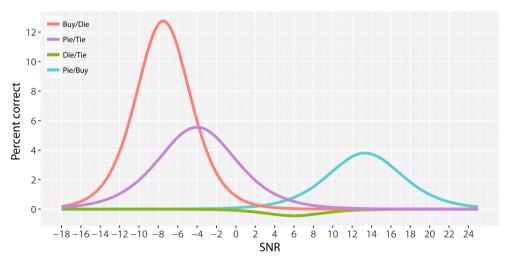


Fig. 5. The distribution of the effect size, as measured by percent correct, is the difference between the passive listening task and the articulatory suppression task for each word pair. The maximum difference observed

was 12.7% for the Buy/Die condition at -7.5 dB, 5.6% for the Pie/Tie condition at -4.0 dB, -.4% for the Die/Tie condition at 6.0 dB, and 3.8% for the Pie/Buy condition at 13.3 dB



are at odds with that explanation given that we did not observe changes in the psychometric function slope as a function of task, suggesting instead that the speech motor system performs a genuine yet modulatory (given the very small effect size) role in speech perception. This is, in fact, consistent with another of our recent reports showing that patients with Broca's area lesions are very mildly impaired on auditory syllable discrimination - measured in terms of perceptual sensitivity, d' - relative to control patients with left mesial temporal-occipital lesions (d' = 4.98 vs. d' = 4.18; Hickok, Costanzo, Capasso, and Miceli, 2011). This corresponds to a difference of about 5% correct separating ceiling from nearceiling performance. Thus, data from multiple experimental modalities (behavioral, stimulation, lesion) appear to converge toward the same conclusion: the motor system *does* play a role in speech perception, but this role is so small as to be nearly undetectable in a laboratory setting.

Potential caveats

There are several caveats regarding the current data that should be noted. First, as described above, performance was somewhat variable across listeners and task conditions (and also across minimal pair conditions, but this was expected), which places us in a poor position to detect small effects with a high degree of certainty. Indeed, as can be inferred from Table 1, the pairwise difference between articulatory suppression and passive listening was, at best, marginally detectable for any given minimal phoneme pair. This difference only reached statistical significance when we collapsed across minimal pairs in a post-hoc analysis. Moreover, we did not maintain strict control over the Type I error by correcting for multiple tests across the post-hoc ANOVAs. For these reasons, the results should be interpreted with caution. However, confidence in the result should be strengthened by the noted convergence with existing data.

Second, articulatory suppression has been shown to induce auditory activation as a consequence of subvocal repetition. Therefore, effects of motor suppression may be conflated with "auditory imagery" resulting in greater disruption of speech perception than we should expect from "pure" speech-motor effects. Indeed, we and others have demonstrated that activity in posterior-auditory brain regions is modulated by covert and overt speech production (cf. Hickok & Poeppel, 2007; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Okada, Matchin, & Hickok, 2018). However, our interpretation of these findings is not only that production activates auditory representations by efference copy, but that such activation reflects a fundamental component of speech production circuits in which internal motor commands are compared to auditory targets for error correction during online speech motor control (Hickok, 2012a). Therefore, the auditory effects generated by silent speech production are a necessary component of any motor speech task and perhaps should not be dissociated from strictly 'motor' effects when the goal is to examine the effect of disrupting motor speech circuits using a behavioral manipulation. Under this holistic interpretation, where articulatory suppression reflects the combined influence of auditory and motor speech production circuits, our result can be interpreted as an upper limit on the purely motor effects of articulatory suppression.

Third, we did not detect reliable differences between articulatory suppression and the other, non-speech dualtask control conditions. Our omnibus statistical test comparing the effect of experimental task across the four tested minimal phoneme pairs revealed a significant two-way interaction, wherein the simple main effect of task was significant only for the Die/Tie pair. Performance was actually best in the articulatory suppression condition for Die/Tie, i.e., performance patterned in the opposite direction as that observed across the other minimal pairs. For Die/Tie, the effect of task was primarily driven by relatively poor performance in the mandible movement and foot tapping conditions, which generally tended to produce poorer performance than passive listening but less consistently so than articulatory suppression. Thus, we did not detect significant differences between mandible movement or foot tapping and passive listening in posthoc analysis, whereas a difference with passive listening was detected for articulatory suppression. Another possible explanation for the pattern of performance in the Die/ Tie condition is that effects of articulatory suppression may be specific to place of articulation distinctions (Buy/Die, Pie/Tie), consistent with recent reports that the place of articulation of speech sounds in decoded in motor areas (Archila-Meléndez et al., 2018; Correia, Jansma, & Bonte, 2015). However, while this would explain the rather anomalous pattern for the Die/Tie condition, the Pie/ Buy condition had a similar pattern of effects as the two place of articulation conditions.

Fourth, we should note that the threshold shift produced by articulatory suppression could, under some circumstances, have a central or cognitive origin. That is, to the extent that any source of error is constant across trials and SNRs, it will produce a change in threshold rather than slope (Fechner, 1860; Morgan et al., 2012). Therefore, if some cognitive mechanism is necessary for the performance of phoneme identification at any and all SNRs, suppression of that mechanism will produce a decrement in performance across SNRs (i.e., a rightward shift of the psychometric function). Phonological working memory is a potential candidate for such a mechanism given the necessity to maintain the results of sensory analysis in memory for long enough to map out and



generate an explicit response in laboratory phoneme identification (Hickok, 2010).

Future directions and conclusions

Despite the noted caveats, we prefer to maintain the most favorable interpretation of the present data with respect to the motor theory of speech perception: together with existing data, this study supports the "weaker" version of the motor theory in which the speech motor system plays a modulatory role in speech perception that scales up with increasing listening difficulty (e.g., in noise). It is worth noting, again, that this "weaker" version is a far cry from the original motor theory, which states unequivocally that the speech motor system is necessary for speech perception (Liberman et al., 1967). In the present study, we were able to detect, at best, a 1-2 dB effect (~4-13% correct) of motor suppression on phoneme identification under ideal test conditions (i.e., a laboratory setting, a large sample of listeners, a large number of trials, and multiple phoneme pairs encompassing both place and voicing distinctions). This finding certainly does not support the original motor theory and, from our perspective, is not sufficiently meaningful to warrant further exploration of the "weaker" version of the motor theory. Given that the basic auditory mechanisms involved in speech perception, and their neural instantiations, are still not well understood (Hamilton, Edwards, & Chang, 2018), further investigation of a motor mechanism that accounts for a very small proportion of the variance in speech perception performance - under limited conditions - is unlikely to advance mechanistic theories of speech perception. Indeed, as we have argued previously, compare the small effect of speech motor disruption - via behavior, stimulation, lesion, or even complete left hemisphere anesthesia (Hickok et al., 2008) - to the catastrophic effect of lesioning the bilateral temporal lobe systems involved in processing speech: pure word deafness, a total loss of the ability to perceive and understand speech under any circumstances (cf., Hickok & Poeppel, 2000, 2004, 2007).

With that said, there remain several open questions with respect to motor system involvement in speech perception. First, it is not known whether the effects of motor suppression observed in laboratory tasks with isolated speech sounds will extend to more naturalistic listening conditions. While one might assume that the effect of motor suppression would increase under more complex listening conditions (e.g., multiple talkers, real-world noise, continuous speech), recent work suggests that the slope of the psychometric function for speech intelligibility is greatly reduced under such conditions (MacPherson & Akeroyd, 2014). This means that an even larger shift in the psychometric function threshold would be required to produce a meaningful effect of motor suppression in terms of changes in intelligibility (percent correct).

Another open question concerns the nature of the neurocomputational mechanisms involved in motor facilitation of speech perception in noise. Some rather compelling theories have been posited based on motor prediction and/or analysis-by-synthesis (Bever & Poeppel, 2010; Liebenthal & Möttönen, 2018; Morillon, Hackett, Kajikawa, & Schroeder, 2015), including instantiation of such mechanisms within computational models (cf. Barnaud et al., 2018; Skipper, Devlin, & Lametti, 2017), but the bulk of the experimental data on cortical motor involvement in speech perception remains correlational (Hickok, 2010; Venezia & Hickok, 2009). A possible direction forward is motivated by two recent neuroimaging and direct cortical recording studies, which converge to suggest that cortical motor representations of heard speech are organized in terms of auditory rather than articulatory features (Arsenault & Buchsbaum, 2016; Cheung, Hamilton, Johnson, & Chang, 2016). This "auditory mode" of processing in the speech motor cortex merits further investigation.

A final open question is related to the emerging translational field of cognitive hearing science, which focuses on the interaction of bottom-up and top-down processes in hearing, particularly in the context of speech understanding in older and hearing-impaired listeners under challenging listening conditions (Arlinger, Lunner, Lyxell, & Kathleen Pichora-Fuller, 2009; Panouillères & Möttönen, 2018; Rönnberg, Rudner, & Lunner, 2011). An active area of research in cognitive hearing science is so-called 'listening effort,' which refers to the notion that degraded auditory processing in older listeners leads to greater recruitment of cognitive resources and thus more effortful listening - even when speech is fully intelligible (Pichora-Fuller et al., 2016). An important goal of compensatory treatments for hearing loss is to mitigate increased listening effort to prevent fatigue and social isolation that may ultimately accelerate cognitive decline (Lin et al., 2013). This is an area where the relatively minor contributions of the speech motor system to perception may be greatly magnified, at least in term of significance. Indeed, a recent study shows that increased frontal speech-motor activity in older adults (relative to young controls) correlates with improved speech sound discrimination in noise and underlies a sharpening of neural phoneme representations (Du, Buchsbaum, Grady, & Alain, 2016). The notion that older listeners may transition toward increased reliance on motor mechanisms merits further investigation, and suggests the speech motor system may be a good target for therapies designed to reduce listening effort under adverse conditions. Stimulation-based motor suppression studies showing effects on response time rather than accuracy (D'Ausilio et al., 2009; Sato, Tremblay, & Gracco, 2009; Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2014) further suggest that the motor mechanism is relevant in the context of listening effort as opposed to



speech perception *per se*, although this may be also be a reflection of the methodological limitations of non-invasive stimulation techniques such as TMS (Devlin & Watkins, 2006; Möttönen & Watkins, 2012).

While all of these questions are interesting and potentially worthwhile to pursue, they focus on a system that is essentially modulatory in the context of speech perception. Their answers are very unlikely to advance our understanding of the fundamental (i.e., auditory) mechanisms underlying speech perception. For this reason, we are hesitant to refer to the "weaker" version of the motor theory as a theory of speech perception at all. Instead, it is a rather simplistic characterization of a class of sensorimotor interactions that: (a) are observable under a limited set of circumstances in the context of speech; (b) are likely not specific to speech (Morillon et al., 2015); and (c) may even be harmful to speech perception in some cases (Hickok, 2014; Whitford et al., 2017). The straightest path toward understanding human speech processing from any perspective - sensory, motor, or sensorimotor - is to put the motor theory of speech perception to bed for good.

References

- Alwan, A., Jiang, J., & Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. Speech Communication, 53(2), 195-209.
- Archila-Meléndez, M. E., Valente, G., Correia, J., Rouhl, R. P., van Kranen-Mastenbroek, V. H., & Jansma, B. M. (2018). Sensorimotor representation of speech perception-cross-decoding of place of articulation features during selective attention to syllables in 7t fmri. eNeuro, ENEURO-0252.
- Arlinger, S., Lunner, T., Lyxell, B., & Kathleen Pichora-Fuller, M. (2009). The emergence of cognitive hearing science. Scandinavian Journal of Psychology, 50(5), 371-384.
- Arsenault, J. S., & Buchsbaum, B. R. (2016). No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. Psychonomic Bulletin & Review, 23(4), 1231-1240
- Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. The Quarterly Journal of Experimental Psychology, 36(2), 233-252.
- Barnaud, M.-L., Bessière, P., Diard, J., & Schwartz, J.-L. (2018). Reanalyzing neurocognitive data on the role of the motor system in speech perception within cosmo, a bayesian perceptuo-motor model of speech communication. Brain and Language, 187, 19-32.
- Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-) emerging program of research for language and vision. Biolinguistics, 4(2-3), 174-200.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex, 10(5), 512-528.
- Bishop, D., Brown, B. B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. Journal of Speech, Language, and Hearing Research, 33(2), 210-219.

- Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [computer program].
- Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433-436.
- Buss, E., Hall III, J. W., & Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults. The Journal of the Acoustical Society of America, 125(2), 1050-1058.
- Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. Elife, 5, e12577.
- Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. The Journal of the Acoustical Society of America, 64(1), 44-56.
- Correia, J. M., Jansma, B. M., & Bonte, M. (2015). Decoding articulatory features from fmri responses in dorsal speech regions. Journal of Neuroscience, 35(45), 15015-15025.
- Craighero, L., Metta, G., Sandini, G., & Fadiga, L. (2007). The mirror-neurons system: data and models. Progress in Brain Research, 164, 39-59.
- D'Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-doppler ultrasound imaging. Philosophical Transactions of the Royal Society B, 369(1644) 20130418.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. Current Biology, 19(5), 381-385.
- Devlin, J. T., & Watkins, K. E. (2006). Stimulating language: insights from tms. Brain, 130(3), 610-622.
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. Nature Communications, 7, 12241.
- D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. Brain and Language, 118(1), 9-14.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a tms study. European Journal of Neuroscience, 15(2), 399-402.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175-191.
- Fechner, G. T. (1860). Elemente der psychophysik: Zweiter theil. Breitkopf und Härtel.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychological Review, 105(4), 611.
- Hamilton, L. S., Edwards, E., & Chang, E. F. (2018). A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Current Biology*.
- Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. Psychonomic Bulletin & Review, 10(2), 435-444.
- Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. Language and Cognitive Processes, 25(6), 749-776.
- Hickok, G. (2012a). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135.
- Hickok, G. (2012b). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. Journal of Communication Disorders, 45(6), 393-402.



- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. Language, Cognition and Neuroscience, 29(1), 2-20.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131-138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of broca's area in speech perception: Evidence from aphasia revisited. *Brain and Language*, 119(3), 214-220.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, 69(3), 407-422.
- Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., ... Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: evidence from wada procedures. Brain and Language, 107(3), 179-184.
- Hillis, A. E. (2007). Aphasia progress in the last quarter of a century. Neurology, 69(2), 200-213.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an meg study. Journal of Cognitive Neuroscience, 14(8), 1125-1138.
- Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Perception and Production of Fluent Speech*, 243–288.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. Attention, Perception, & Psychophysics, 63(8), 1421-1455.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. Perception, 36(14),
- Kontsevich, L. L., & Tyler, C. W. (1999). Distraction of attention and the slope of the psychometric function. JOSA A, 16(2), 217-222.
- Kuhl, P. K., & Miller, J. D. (1971). Speech perception by the chinchilla: Voiced-voiceless distinction in. Annals of the New York Academy of Sciences, 185, 345.
- Laurent, R., Barnaud, M.-L., Schwartz, J.-L., Bessière, P., & Diard, J. (2017). The complementary roles of auditory and motor information evaluated in a bayesian perceptuo-motor model of speech perception. *Psychological Review*.
- Leek, M. R., Hanna, T. E., & Marshall, L. (1992). Estimation of psychometric functions from adaptive tracking procedures. Perception & Psychophysics, 51(3), 247-256.
- Lenneberg, E. H. (1962). Understanding language without ability to speak: a case report.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74(6), 431.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.
- Liebenthal, E., & Möttönen, R. (2018). An interactive model of auditorymotor speech perception. Brain and Language, 187, 33-40.
- Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., ... et al. (2013). Hearing loss and cognitive decline in older adults. JAMA Internal Medicine, 173(4), 293-299.
- Liu, H. T., Squires, B., & Liu, C. J. (2016). Articulatory suppression effects on short-term memory of signed digits and lexical items in hearing bimodal-bilingual adults. Journal of Deaf Studies and Deaf Education, 21(4), 362-372.

- MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, 18.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. Current Biology, 17(19), 1692-1696.
- Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia pathologic and clinical. Neurology, 28(4), 311-311.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). Reason, 4(2), 61-64.
- Morgan, M., Dillenburger, B., Raphael, S., & Solomon, J. A. (2012).
 Observers can voluntarily shift their psychometric functions without losing sensitivity. Attention, Perception, & Psychophysics, 74(1), 185-193
- Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. Current Opinion in Neurobiology, 31, 230-238.
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. Journal of Neuroscience, 29(31), 9819-9825.
- Möttönen, R., & Watkins, K. E. (2012). Using tms to study the role of the articulatory motor system in speech perception. Aphasiology, 26(9), 1103-1118.
- Okada, K., Matchin, W., & Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. Psychonomic Bulletin & Review, 25(1), 423-430.
- Panouillères, M. T., Boyles, R., Chesters, J., Watkins, K. E., & Möttönen, R. (2018). Facilitation of motor excitability during listening to spoken sentences is not modulated by noise or semantic coherence. Cortex, 103, 44-54.
- Panouillères, M. T., & Möttönen, R. (2018). Decline of auditory-motor speech processing in older adults with hearing loss. Neurobiology of Aging, 72, 89-97.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437-442.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). Ear and Hearing, 37, 5S-27S.
- Potvin, P. J., & Schutz, R. W. (2000). Statistical power for the two-factor repeated measures anova. Behavior Research Methods, Instruments, & Computers, 32(2), 347-356.
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. Journal of Anatomy, 197(3), 335-359.
- Prins, N., & Kingdon, F. (2009). Palamedes: Matlab routines for analyzing psychophysical data. Palamedes: matlab routines for analyzing psychophysical data.
- Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? evidence from five cases with damage to the purported human mirror system. Neurocase, 17(2), 178-187.
- Rönnberg, J., Rudner, M., & Lunner, T. (2011). Cognitive hearing science: The legacy of stuart gatehouse. Trends in Amplification, 15(3), 140-148.
- Saeki, E., & Saito, S. (2004). Effect of articulatory suppression on taskswitching performance: Implications for models of working memory. Memory, 12(3), 257-271.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. Cognitive Brain Research, 23(2), 429-435.
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. Brain and Language, 111(1), 1-7.
- Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2014). Causal influence of articulatory motor cortex on



- comprehending single spoken words: Tms evidence. Cerebral Cortex, 25(10), 3894-3902.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. Brain and Language, 164, 77-105.
- Venezia, J. H., & Hickok, G. (2009). Mirror neurons, the motor system and language: from the motor theory to embodied cognition and beyond. Language and Linguistics Compass, 3(6), 1403-1416.
- Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. Frontiers in Psychology, 3, 157.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41(8), 989-994.
- Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. Trends in Cognitive Sciences, 9(11), 519-527.

- Whitford, T. J., Jack, B. N., Pearson, D., Griffiths, O., Luque, D., Harris, A. W., ... Le Pelley, M. E. (2017). Neurophysiological evidence of efference copies to inner speech. *Elife*, 6.
- Wilson, S. M. (2009). Speech perception when the motor system is compromised. Trends in Cognitive Sciences, 13(8), 329.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nature Neuroscience, 7(7), 701-702.
- Wu, Z.-M., Chen, M.-L., Wu, X.-H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. Neuroscience Bulletin, 30(3), 490-496.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

