frontiers in Communication

ORIGINAL RESEARCH
published: 27 May 2020
doi: 10.3389/fcomm.2020.00034

Speech Perception Triggers Articulatory Action: Evidence From Mechanical Stimulation

Iris Berent[1]*, *Melanie Platt*[1], *Rachel Theodore*[2], *Evan Balaban*[3], *Peter J. Fried*[4,5] and *Alvaro Pascual-Leone*[5,6,7]

[1] Department of Psychology, Northeastern University, Boston, MA, United States, [2] Department of Speech, Language, and Hearing Sciences, University of Connecticut, Mansfield, MA, United States, [3] Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Getafe, Spain, [4] Berenson-Allen Center for Noninvasive Brain Stimulation, Beth Israel Deaconess Medical Center, Boston, MA, United States, [5] Department of Neurology, Harvard Medical School, Boston, MA, United States, [6] Department of Neurology, Hinda and Arthur Marcus Institute for Aging Research and Center for Memory Health, Hebrew SeniorLife, Boston, MA, United States, [7] Guttmann Brain Health Institute, Institut Guttmann de Neurorehabilitació, Universitat Autonoma de Barcelona, Barcelona, Spain

Neuroimaging studies show that the brain articulatory motor system is activated during speech perception. These results, however, cannot determine whether the motor system is engaged in motor simulation (e.g., lip movement) or in non-motor computations. To gauge the causal role of the articulatory motor system, here we examine the effect of mechanical stimulation. Participants heard speech tokens, ambiguous with respect to voicing—either labial (*ba/pa*) or coronal (*da/ta*)—while lightly biting on either the lips or tongue. Compared to incongruent stimulation (e.g., lips, with *da/ta*), congruent stimulation (e.g., tongue, with *da/ta*) shifted the voicing percepts, and improved discrimination sensitivity (d'). These results demonstrate that adults engage the articulatory motor system in speech perception even when it is irrelevant to response. The convergence with transcranial magnetic stimulation experiments suggests that speech perception engages articulatory action. These conclusions illuminate the links between embodiment and cognition.

Keywords: speech perception, embodiment, motor simulation, phonetic categorization, voicing, TMS

OPEN ACCESS

Edited by:
Ludovic Ferrand, Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Jean-Luc Schwartz, Centre National de la Recherche Scientifique (CNRS), France
Takenobu Murakami, Fukushima Medical University, Japan

***Correspondence:**
Iris Berent
i.berent@northeastern.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *06 February 2020*
Accepted: *28 April 2020*
Published: *27 May 2020*

Citation:
Berent I, Platt M, Theodore R, Balaban E, Fried PJ and Pascual-Leone A (2020) Speech Perception Triggers Articulatory Action: Evidence From Mechanical Stimulation. Front. Commun. 5:34. doi: 10.3389/fcomm.2020.00034

INTRODUCTION

Speech is the gateway to language. The speech signal, however, is analog and continuous, whereas linguistic categories (e.g., /b/ vs. /p/) are discrete. Accordingly, to enter the language system, speech must first be classified into discrete categories (Liberman et al., 1967). How phonetic categories are extracted has been the subject of an ongoing controversy.

One possibility is that the perception of phonetic categories is principally informed by acoustic cues (e.g., Holt and Lotto, 2008). An alternative account links speech perception to action: to categorize speech sounds, a listener must simulate their production (Liberman et al., 1967; Fadiga et al., 2002; for review Skipper et al., 2017). In its strongest form, this hypothesis predicts that, to identify labial sounds, such as *b* and *p*, the listener must enact the production of such sounds by her lips.

Numerous studies have indeed demonstrated that speech perception engages the brain articulatory motor areas (e.g., Pulvermüller et al., 2006; for review Badino et al., 2014; Skipper et al., 2017), and their disruption by transcranial magnetic stimulation (TMS) alters the identification of

congruent speech sounds (e.g., Smalle et al., 2014). Such TMS results are significant because they suggest that articulatory brain motor areas have a causal role in phonetic categorization. Many researchers have further taken these findings as evidence that speech perception requires motor ***simulation*** (Fadiga et al., 2002; Pulvermüller et al., 2006). But this conclusion is far from certain.

One concern is that, in most TMS studies (D'Ausilio et al., 2009, 2012; Möttonen and Watkins, 2009; Smalle et al., 2014) the stimulated brain area (e.g., the lip motor area) was relevant to only one of the responses (e.g., to *ba,* but not *da).* Accordingly, the effect of stimulation could be due to the role of the lip motor areas in the selection of *ba* response, rather than in speech perception.

Somatosensory brain areas, moreover, have been implicated in multiple functions. "Visual" brain areas are engaged by tactile (Kauffman et al., 2002) and auditory (Amedi et al., 2007; Striem-Amit et al., 2012) inputs, and they can even mediate syntactic processing in both blind (Bedny et al., 2011) and sighted individuals (Elli et al., 2019; Pant et al., 2019). Deaf signers, by contrast, are known to engage speech motor areas during the production of signs (Martino et al., 2017). The multiple functionality of the somatosensory brain system suggests caution in evaluating the role of motor brain areas. One possibility is that the activation of articulatory motor areas in speech perception indeed reflects their engagement in motor simulation. But another possibility is that these brain areas are engaged in functions that are entirely unrelated to articulation, such as a purely acoustic analysis, phonetic processing, or even higher order linguistic (e.g., phonological) computations.

To elucidate the functional role of the motor system in speech perception, it is thus desirable to contrast the effects of brain stimulation revealed by TMS studies, with those of mechanical stimulation of the motor end-organ. If hearers activate the lip motor area in the brain during linguistic tasks because it effectively controls the movement of their lips, then speech perception should be likewise modulated by the stimulation of the lips *mechanically* (e.g., by biting on the lips relative to the tongue). This effect, moreover, should be present even when the lips are equally relevant to both responses (e.g., *b* vs. *p*).

Recent results from infants bear out this prediction. When six-month old infants restrain the movement of their tongue tip, they lose their sensitivity to nonnative sounds that engage this articulator (Bruderer et al., 2015). Comparable manipulations with adults, however, have yielded a more complex picture.

Some studies either found no reliable effects (Cooper et al., 1975) or only elicited a response bias (Sato et al., 2011; Venezia et al., 2012). Similarly, dynamic articulatory rehearsal did not alter either the slope of the phonetic identification function (Stokes et al., 2019) or the McGurk effect (Matchin et al., 2014). Other studies with adult participants did find that the identification of speech sounds is modulated by mechanical stimulation (Ito et al., 2009; Nasir and Ostry, 2009; Stokes et al., 2019; Ogane et al., 2020), silent articulation (Sams et al., 2005), mouthing and imagery (Scott et al., 2013), the application of air puff (suggesting aspiration; Gick and Derrick, 2009) and imagery alone (Scott, 2013). These effects, however, were either non-selective (Stokes et al., 2019) or limited to identification choice (Gick and Derrick, 2009; Ito et al., 2009; Nasir and Ostry, 2009; Ogane et al., 2020), so it is unclear whether they reflect change in perceptual sensitivity or response selection alone.

The divergent effects of mechanical stimulation on infants and adults are open to two conflicting interpretations. One possibility is that motor simulation mediates only the *acquisition* of phonetic categories in early development, but not the processing of familiar speech categories by adults. Alternatively, speech simulation might mediate phonetic categorization in both infants and adults, but the divergent results could be due to methodological limitations of the adult studies. Indeed, as in the TMS studies, most previous research on mechanical stimulation with adults has systematically confounded the effect of stimulation (e.g., of the *lips* vs. *tongue*) with response choice (e.g., ba vs. da, Cooper et al., 1975; Sams et al., 2005; Ito et al., 2009; Nasir and Ostry, 2009; Venezia et al., 2012; Scott et al., 2013; Matchin et al., 2014). Since the stimulated articulator (e.g., lips) is only relevant to one of the two responses (*ba*), this manipulation might be particularly vulnerable to response bias. The question thus remains whether motor stimulation could alter phonetic categorization when the role of stimulated articulators in response is minimized.

To date, only one recent study has demonstrated that mechanical stimulation selectively modulates perceptual sensitivity, not merely accuracy (Masapollo and Guenther, 2019). Results showed that perceptual sensitivity (A') to vowel contrasts improves when people perform congruent relevant gestures. Specifically, people became more sensitive to the English /ɛ/–/æ/ contrast (a contrast that concerns the mandibular posture) when participants held in their mouth a block compared to a tube; the tube, in contrast, improved the perception of the French and English /u/ sounds—a contrast defined by the labial posture. No previous study, however, has asked whether mechanical suppression could selectively affect perceptual sensitivity to consonants.

To address this question, the present experiments examine the effect of mechanical stimulation on the perception of labial (*ba/pa*) and coronal (*da/ta*) sounds. Participants heard these sounds while concurrently biting on either their lips or tongue. With this design, the stimulated articulator (e.g., lips) is either relevant to both response options (e.g., for both *ba* and *pa*), or to neither (*da* or *ta*), effectively dealing with the problem of response bias. A control condition simply asked participants to relax while listening to the sounds.

Experiment 1 examined the identification of a single sound, ambiguous between either *ba* and *pa* or *da* and *ta*; Experiments 2–4 elicited discrimination between sound pairs. Of interest is whether the identification and discrimination performance (e.g., on labials) is modulated by the stimulation of a congruent articulator (e.g., the lips) relative to an incongruent articulator (e.g., the tongue).

Although the lips and tongue are not relevant to voicing specifically, their incongruent engagement could disrupt the perception of voicing by interfering with the gestural articulatory plan as a whole. Consequently, the stimulation of the lips and tongue should selectively modulate the perception of congruent speech sounds. In particular, the perception of the labial *ba/pa*

contrast should be more strongly affected by stimulating the lips compared to the tongue, whereas the perception of the *da/ta* contrast should be more strongly affected by stimulation the tongue compared to the lips.

We note that the direction of the congruence effect cannot be predicted a priori, as the stimulation of a congruent articulator (e.g., lips, for *ba/pa*) could either impair speech perception or improve it (by either impeding motor simulation or stimulating it). Additionally, because the stimulation and control conditions differ on multiple levels (e.g., their putative roles in articulatory simulation, their attention and memory demands), we make no a priori predictions concerning the effect of stimulation relative to the control condition. The key question here is thus whether the stimulation of a congruent articulator *differs* in its effects from the stimulation of an incongruent articulator.

EXPERIMENT 1: IDENTIFICATION

Experiment 1 examined the identification of speech sounds that were ambiguous with respect to their voicing. One block of trials elicited the identification of a labial contrast (e.g., *ba* or *pa*?) another featured a coronal contrast (e.g., *da* or *ta*?). Unbeknownst to participants, however, the sounds they heard in each block corresponded to a *single* ambiguous speech token, repeated multiple times. This single ambiguous speech token corresponded to the midpoint of a 10-step voicing continuum (step 6), either a labial or a coronal (hereafter, *ba/pa* and *da/ta*, respectively). Each such block was repeated thrice, and throughout each presentation, participants were instructed to concurrently bite on either their lips, their tongue, or relax (a control condition). Accordingly, each stimulated articulator was congruent with one contrast: the lips were congruent with the labial *ba/pa* contrast, whereas the tongue was congruent with the coronal *da/ta* contrast. Moreover, each articulator (e.g., the lips) was either relevant to both responses (to *ba* and *pa*) or to neither (*da* and *ta*). In this fashion, the effect of stimulation is dissociated from response selection.

If speech perception (e.g., of the *ba/pa* contrast) elicits motor simulation, then the stimulation of a congruent articulator (e.g., of the lips) should selectively modulate phonetic identification relative to the incongruent articulator (e.g., the tongue), and this effect should emerge even when stimulation is disconfounded with response selection.

MATERIALS AND METHODS

Participants

Forty-eight participants took part in the experiment. Participants in this and all subsequent experiments were native English speakers, students at Northeastern University. They took part in the experiment in partial fulfillment of a course credit. Sample size was determined by previous pilot work in our lab, as no previous study has examined the selective effect of mechanical stimulation on adult participants.

The study was approved by the IRB at Northeastern University. All methods were performed in accordance with the relevant guidelines and regulations, and informed consent was signed by all participants or their parents and/or legal guardians (for participants under the age of 18 years).

To ensure that the stimuli in Experiment 1 were perceptually ambiguous, we excluded 10 participants whose performance in the control condition was at floor/ceiling. We defined floor/ceiling as either (a) responses falling 2 SD above/below the means; or (b) responses falling either above 0.9 or below 0.1. We applied rule (b) only if rule (a) yielded impossible exclusion values (below 0 or above 1)[1].

Materials

The critical trials featured two tokens. Each such token was ambiguous with respect to its voicing. One such token was labial (ambiguous between *pa* and *ba*); a second token was coronal (ambiguous between *ta* and *da*). These stimuli were each obtained from a 10-step voicing continuum, generated from a recording of the endpoints by a native English speaker. These stimuli were used in Berent et al. (2016) and their description (below) is adapted from the Methods section therein. Experiment 1 featured the ambiguous midpoint of each continuum (step 6; the procedure for selecting this step is described below).

In addition to the two voicing tokens, Experiment 1 also included the midpoint of a third contrast, ambiguous with respect to its place of articulation (*ba/da*). Results showed that the stimulation of the lip increased the proportion of "*ba*" responses (in line with past TMS results; D'Ausilio et al., 2009, 2012; Möttonen and Watkins, 2009; Smalle et al., 2014). But because the effect of stimulation on the identification of this contrast could be due to response bias, we will not discuss it further (for results, see **Supporting Information**).

The Preparation of the Continua

Stimulus manipulations were performed using SIGNAL software (Engineering Design, Berkeley, CA) and Matlab (Mathworks, Natick, MA). All continua were made from recordings of isolated syllables produced by native speakers (16 bits, 44.1 kHz sampling rate). The *ba-pa* continuum was produced by removing the DC component from both endpoint syllables, setting non-vocalization portions of the recording to zero, and truncating the recording lengths to the shorter of the two stimuli. The *ba* syllable was rotated (zero-value segments added/subtracted at the beginning and end, while keeping the file length the same) so that its vowel periodicity aligned with the *pa* syllable, and a "hybrid *pa*" syllable was created using the first 159.19 ms of the *pa* and the rest of the *ba* from 159.19 ms to the recording end, joined at a zero-crossing. The noisy initial part (the aspiration phase) of this "hybrid *pa*" syllable was then progressively shortened at successive zero crossings occurring within the time interval between 114–131 ms to make the eight steps of the continuum. The "*pa*" endpoint had a value of 57.8 ms from the start of articulation to the end of the aspiration phase; this value was 37.1 ms for the "*ba*" endpoint.

[1]Experiment 1 included two sub-groups of participants ($N = 24$ each). We thus applied the exclusion rules to each group separately. For the first group, we included data falling between 0.2–0.98 and.1–0.84 (for *ba-pa* and *da-ta*, respectively); for the second group, inclusion values were 0.1–0.95 and 0.1–0.77 (for *ba-pa* and *da-ta*, respectively).

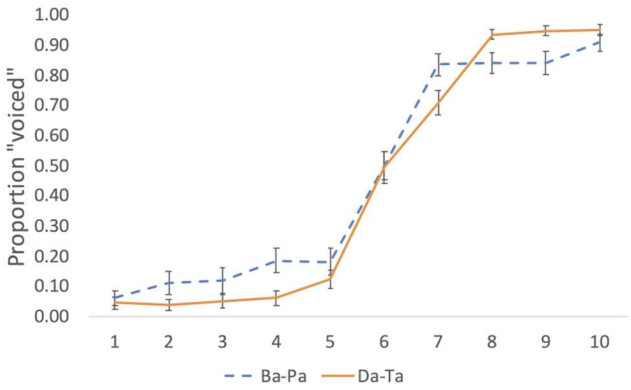

FIGURE 1 | Pilot identification results for the labial (Ba-Pa) and coronal (Da-Ta) voicing continua. Error bar are standard errors of the means.

The *da-ta* continuum was produced by removing the DC component from both endpoint syllables, setting non-vocalization portions of the recording to zero, and truncating the recording lengths to the shorter of the two stimuli. The *da* syllable was then rotated to align its vowel periodicity with the *ta* syllable, and a "hybrid *ta*" syllable was created using the first 183.38 ms of the *ta* and the rest of the *da* from 183.38 ms to the recording end. The first 6 stimuli in the continuum were made by progressively shortening the noisy part of the "hybrid *ta*" syllable starting from the zero crossing that was proximal to the start of the voicing for the vowel, and proceeding backwards at successive zero crossings occurring every 7–10 ms. The rest of the stimuli in the continuum were made by successively replacing the remaining portion of the noisy signal before the start of the voicing by successive voiced vowel periods present in the original *da* syllable at these same positions in time (splicing done at zero crossings), ending up with a perfect reproduction of the original *da* syllable at the end of the continuum.

Midpoint Estimates for the Voicing Continua

We estimated the perceptual midpoint of each continuum by analyzing the results of a pilot experiment. In the experiment, a separate group of 24 native English speakers (students at Northeastern University) identified the same 10 step continua without stimulation. Results are plotted in **Figure 1**.

We calculated the midpoint as follows. For each participant, the quickpsy package in R (Linares and López-Moliner, 2016) was used to fit the identification responses to a cumulative normal function, constraining the guess and lapse rates to be within 5% following procedures for fitting psychometric functions (e.g., Wichmann and Hill, 2001; Clayards et al., 2008), with the mean of the derived function representing the perceptual midpoint. We then selected the mean value across participants in each continuum. Step 6 was the perceptual midpoint in each of the three continua. To render the repetition of the stimulus less conspicuous, trials were separated by a white noise stimulus presented at 68 dB for a duration of 500 ms. The white noise was generated in Praat (Boersma and Weenink, 2018) using a custom-made script.

Design

Each block of trials featured a single contrast (e.g., *ba/pa*) repeatedly, paired with a single stimulation condition. Each such contrast was crossed with each of the three stimulation conditions (lips/tongue/control). Block order was counterbalanced across participants. Each stimulus was repeated 60 times with each of the three stimulation conditions, for a total of 540 trials (3 contrasts × 3 stimulation × 60 repetitions).

Procedure

Each trial began with a screen, displaying a fixation point (*) and the trial number. Participants initiated the trial by pressing the space bar. Their response triggered the presentation of a single auditory stimulus, which participants were asked to identify (e.g., *da* or *ta*) by pressing one of two keys. Their response triggered the presentation of white noise for 500 ms, which was replaced by the next trial. Participants had up to 4 s to respond; responses slower than 2,500 ms received a "too slow" message. Response time in this and all subsequent experiments is reported from the onset of the auditory stimulus.

Participants performed the identification task while either lightly biting on their lips, lightly biting on their tongue, or "doing nothing" (control). Participants were instructed to attempt to only engage the relevant articulator and leave the other free to move. The three stimulation conditions (lips, tongue, control) were presented in three blocks, counterbalanced for order. Each such block further featured the three ambiguous stimuli in three separate sub-blocks, counterbalanced for order. To remind participants of the stimulation task, in each block, participants were given a printed picture describing the relevant stimulation condition (see **Figure 2**). This picture was displayed on the computer screen only at the beginning of each block; as a reminder, we also placed a hardcopy of the picture on the computer desk, next to the participant. Throughout the experiment, however, participants attended primarily to the computer screen[2].

Data Analysis

In this and all subsequent experiments, the analyses of response time eliminated outliers (correct responses falling 2.5 SD above or below the mean), which amounted to <3.5% of the data in each experiment.

Experiment 1 was analyzed in *R*. Individual trial responses (0 = voiceless, 1 = voiced) were fit to a generalized linear mixed-effects model with the binomial response family. The fixed effects were contrast-coded and included contrast (*ba/pa* = −1/2, *da/ta* = 1/2) and stimulation; stimulation was entered into the model as two orthogonal contrasts, one for lips vs. tongue (control = 0, lips = −1/2, tongue = 1/2) and one for control

[2]It is unlikely that the presentation of this picture can explain our results, as (a) participants attended primarily to the computer screen, not the image placed on the desk; and (b) Rosenblum and Saldaña (1996) have shown that a static facial image is insufficient to elicit audiovisual integration (as in the McGurk effect, McGurk and MacDonald, 1976) even when participants attended to the image.

FIGURE 2 | An illustration of the lip stimulation, tongue stimulation and the control condition.

vs. stimulation (control = −2/3, lips = 1/3, tongue = 1/3); all models included random intercepts by subject and random slopes for both contrast and stimulation by subject. For response time (RT), trial-level response times for voiced responses were fit to a linear mixed-effects model with the same fixed and random effects structure as for the identification model.

RESULTS

Figure 3 presents the proportion of "voiced" responses and the response time to select the "voiced" option. An inspection of the means suggests that the stimulation of the congruent articulator facilitated the voicing response. For the coronal *da/ta* contrast, tongue stimulation increased the proportion of "voiced" responses, whereas for the labial *ba/pa* contrast, lips stimulation sped up response time.

A Comparison of the Lips/Tongue/Control Conditions

To evaluate the reliability of these observations, identification and response time were analyzed in separate mixed-effects models (for full model description, see methods).

Results yielded a significant main effect of contrast ($b = 0.51$, $SE = 0.14$, $Z = 3.64$, $p < 0.0004$) as the labial continuum yielded a lower proportion of "voiced" responses. Critically, the effect of lips vs. tongue stimulation interacted with the contrast type (*ba/pa* and *da/ta*; $b = 0.37$, $SE = 0.09$, $Z = 4.03$, $p < 0.003$). This same interaction was also significant in the analysis conducted over "voiced" response time ($b = -80.54$, $SE = 13.43$, $t = -6.00$, $p < 0.0001$).

The analysis of response time also yielded several additional non-selective effects of contrast and stimulation. We found that lip stimulation sped up the overall "voiced" response compared to tongue simulation ($b = 39.18$, $SE = 18.58$, $t = 2.08$, $p < 0.05$), and that the control condition resulted in overall slower responses relative to lips/tongue stimulation ($b = -49.54$, $SE = 15.34$, $t = -3.23$, $p < 0.003$). We also found that, when compared to the lips/tongue conditions, the control conditions yielded slower "voiced" responses in the *da/ta* relative to the *ba/pa* continuum ($b = -59.75$, $SE = 11.49$, $t = -5.20$, $p < 0.0001$). Since these two control conditions did not differ in the absence of stimulation (see **Figure 1**), the observed difference must be due to the stimulation context. As noted, differences between the stimulation and control conditions could reflect multiple sources, unrelated to motor simulation (e.g., attention demands). Accordingly, our analysis proceeded to contrast the effect of stimulation by directly comparing the two stimulation conditions (lips vs. tongue) for the two contrasts (*ba/pa* vs. *da/ta*).

A Comparison of Lips vs. Tongue Stimulation

The interaction (2 stimulation × 2 contrast) was significant in both response selection ($b = 0.41$, $SE = 0.09$, $Z = 4.41$, $p < 0.0001$) and response time ($b = -75.49$, $SE = 13.23$, $t = -5.70$, $p < 0.0001$). Additionally, we found a significant effect of contrast in response selection ($b = 0.57$, $SE = 0.22$, $Z = 2.65$, $p < 0.009$)[3].

To interpret these interactions, we thus proceeded to compare the effect of lips and tongue stimulation in each continuum separately. For the *da/ta* contrast, the stimulation of the congruent articulator (tongue) increased the proportion of "voiced" responses ($b = 0.43$, $SE = 0.21$, $Z = 2.07$, $p < 0.04$; for response time: $t < 1$). For *ba/pa*, stimulation of the congruent articulator (lips) decreased response time to the "voiced" option ($b = 72.58$, $SE = 29.00$, $t = 2.50$, $p < 0.02$); for response selection, $Z < 1$). Thus, in each continuum, congruent stimulation reliably facilitated the voicing response relative to the incongruent condition.

EXPERIMENTS 2-4: DISCRIMINATION

To further dissociate the effect of stimulation from response selection, Experiments 2–4 examined the effect of mechanical stimulation on discrimination. In each trial, people heard two sounds from a single continuum. Experiments 2–3 contrasted a stimulus with ambiguous voicing (the midpoint of the continuum, step 6) with a voiceless step—either step 2 (in

[3]A parallel analysis, examining the effect of stimulation on response time for "voiceless" responses yielded similar results. The omnibus comparison of the three suppression conditions (3 suppression × 2 contrast) yielded a reliable interaction ($b-38.264$, $SE = 13.077$, $t = -2.926$, $p < 0.004$). A follow up comparison of the effect of Lips vs. Tongue stimulation (2 stimulation × 2 contrast) further yielded a significant interaction ($b = -40.240$, $SE = 13.169$, $t = -3.056$, $p < 0.003$). The simple effect of stimulation (Lips vs. Tongue) was significant for the *ba/pa* ($b = 62.909$, $SE = 29.613$, $t = 2.124$, $p = 0.04$) but not for *da/ta* contrast ($b = 33.964$, $SE = 20.490$, $t = 1.658$, $p = 0.11$).

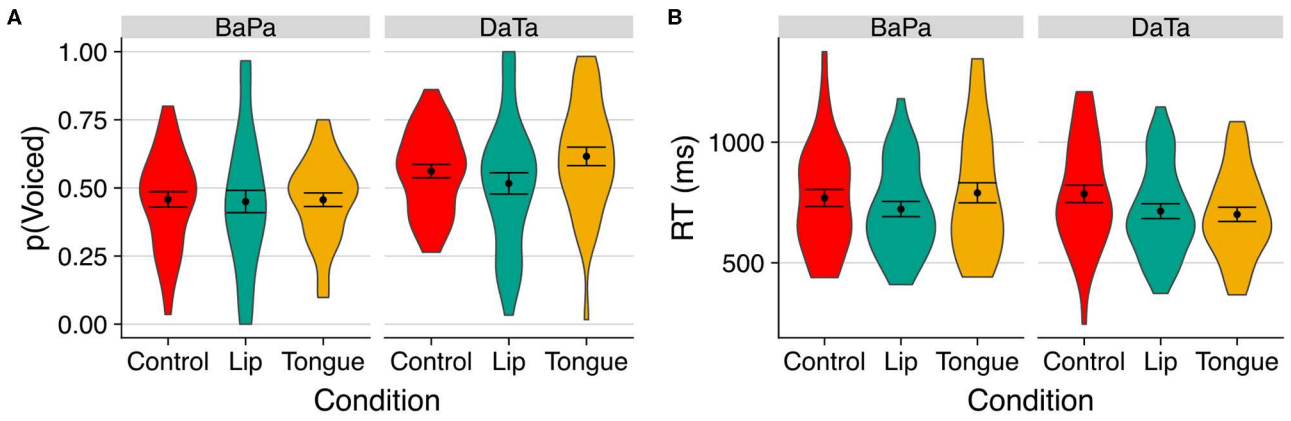

FIGURE 3 | The effect of stimulation on the perception of voicing of labial and coronal sounds in Experiment 1. **(A)** indicates proportion of *voiced* responses; **(B)** indicates response time to the *voiced* option. Error bars are standard error of the mean.

Experiment 2), or a step selected to yield the same proportion of voiced responses by each articulator (in Experiment 3). Experiment 4 contrasted two voiceless steps (steps 2 and 4). Of interest is whether the stimulation of a congruent articulator modulates perceptual sensitivity (as measured by d').

METHODS

Participants

Experiments 2–4 (discrimination) each employed a distinct group of 24 participants.

Materials

Each trial contrasted two tokens (hereafter, tokens *a* and *b*), sampled from one of the two stop continua (labial or coronal) described in Experiment 1. Half of the trials presented identical pairs of tokens (*a-a*; *b-b*) and half presented non-identical pairs (*a-b; b-a*).

Experiments 2–3 contrasted the midpoint with a voiceless step. In Experiment 2, this voiceless token corresponded to step 2 of each continuum; in Experiment 3, we selected the voiceless step to yield comparable (∼0.12) voicing responses in the labial and coronal continua (for results, see **Figure 1**). The steps that yielded this value differed in the labial and coronal contrasts, and they corresponded to steps 3 and 5, respectively. Finally, Experiment 4 contrasted steps 2 and 4 of the two continua.

Design

As in Experiment 1, the voicing contrast of each place of articulation (labial vs. coronal) was presented in a separate block, crossed with each of the three stimulation conditions (lips/tongue/control), with block order counterbalanced across participants.

Each such block contrasted two tokens of the same phoneme (token *a* and *b*), combined in four pairs (*a,a; b,b; a,b; b,a*). Each such pair was repeated 15 times in each of the three stimulation conditions for a total of 360 trials (4 tokens × 15 repetitions × 2 contrasts × 3 stimulation). Trial order within a block was randomized.

Procedure

Each trial began with a screen indicating a fixation point (*) and the trial number. Participants pressed the space bar to initiate the trial. Their response triggered the presentation of the first auditory token (for 1,000 ms) followed by a white noise (for 1,200 ms) and second stimulus (for 600 ms). Participants were given up to 2,500 ms to respond. Participants performed the task while lightly biting on either their lips, biting on their tongue, or relaxing, as described above.

RESULTS

Data from Experiment 2 (contrasting the ambiguous step 6 with the relatively voiceless step 2) are shown in **Figure 4**. An inspection of sensitivity (d'; hits are defined as correct "identical" responses; false alarms are incorrect responses to "nonidentical" trials) suggested that for the labial contrast, discrimination was harder than the coronal contrast, and it was not selectively affected by stimulation. In contrast, for the coronal stimuli, tongue stimulation selectively improved discrimination (d') at the cost of a slower response time.

This interpretation was supported by a 2 contrast × 3 stimulation ANOVA. The analysis of d' yielded a reliable main effect of contrast [$F_{(1,23)} = 23.20$, $p < 0.0001$, $\eta^2 = 0.33$], as well as a significant contrast × stimulation interaction [$F_{(2,46)} = 4.70$, $p < 0.02$, $\eta^2 = 0.02$]. The same interaction was also significant in the analysis of response time to identity trials [i.e., *a,a; b,b;* $F(F_{(2,46)} = 4.17$, $p < 0.03$, $\eta^2 = 0.03$]. No significant effects obtained for nonidentity trials (all <1.4, $p > 0.26$). A follow up (2 contrast × 3 stimulation × 2 step) ANOVA confirmed that the effect of stimulation was not further modulated by step ($p > 0.17$).

We next probed the effect of stimulation in the two continua using planned contrasts. For labial sounds, we found no selective effects of stimulation. Although, when compared to the control

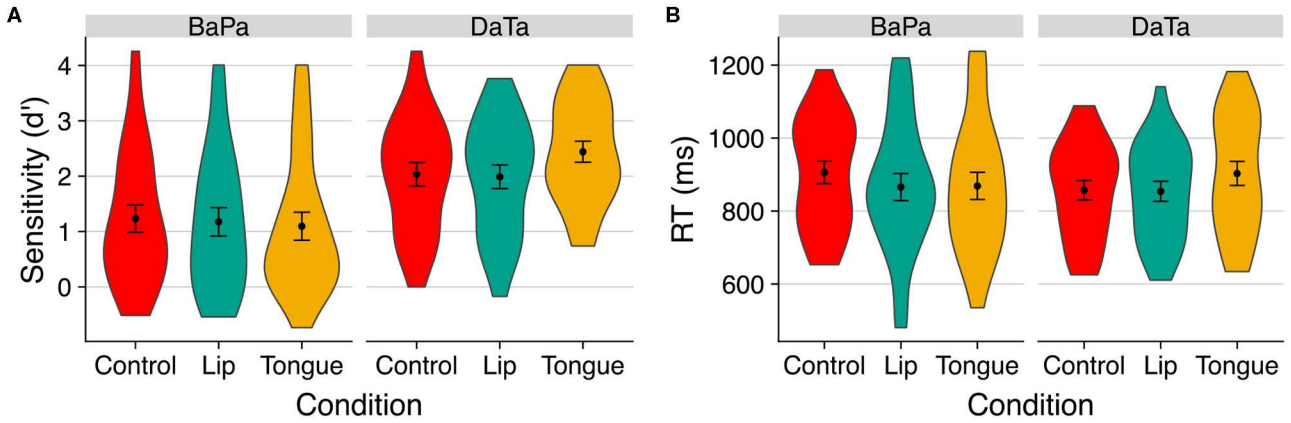

FIGURE 4 | The effect of stimulation on **(A)** sensitivity (d') and **(B)** response time to identity trials in Experiment 2. Error bars are standard error of the mean.

condition, lips [$t_{(46)} = 1.97$, $p < 0.06$] and tongue [$t_{(46)} = 1.83$, $p < 0.08$] stimulation each produced a marginally significant decrease in response time to labial sounds, lips and tongue stimulation did not reliably differ ($t < 1$). There were likewise no effects of stimulation on sensitivity to labial sounds (all $t < 1$).

Stimulation, however, selectively modulated sensitivity to coronal sounds. Here, tongue stimulation improved sensitivity relative to both the lips [$t_{(46)} = 3.17$, $p < 0.003$] and control [$t_{(46)} = 2.84$, $p < 0.007$] conditions. Tongue stimulation also reliably increased response time relative to both the lips [$t_{(46)} = 2.41$, $p < 0.02$] and control [$t_{(46)} = 2.25$, $p < 0.03$] conditions. Thus, in Experiment 2, the stimulation of the congruent articulator (tongue) improved sensitivity to coronal sounds at the cost of slower response time. The lack of stimulation effects for labial sounds may have been due to their lower discriminability relative to the coronal stimuli (in pilot results, step 2 of the labial contrast elicited a higher proportion of "voiced" responses ($M = 0.11$) relative to of the coronal contrast ($M = 0.04$); see **Figure 1**).

To address this possibility, in Experiment 3 we contrasted the midpoint (step 6) with a voiceless step that was selected to yield similar voicing responses in the labial and coronal continua (∼0.12). The continuum steps that corresponded to this target value differed for the labial and coronal stimuli (steps 3 and 5, respectively, hereafter the "voiceless" step).

Results (see **Figure 5**) revealed that coronal stimuli yielded faster responses under the congruent stimulation, whereas stimulation did not selectively modulate response to labial sounds.

We first examined the effect of stimulation via a 2 contrast × 3 stimulation ANOVA of sensitivity (d'). Results only yielded a reliable effect of contrast [$F_{(1,23)} = 8.80$, $p < 0.007$, $\eta^2 = 0.18$], as labial sounds produced higher sensitivity. No other effect was significant (all $p > 0.3$).

We next turned to examine the effect of stimulation on response time using a 2 contrast × 3 stimulation × 2 step (the midpoint vs. voiceless step) ANOVA, applied separately for identity and nonidentity trials. No other effects were significant (all $p > 0.27$). For identity trials, however, there was a significant effect of step [$F_{(1,23)} = 8.36$, $p < 0.009$, $\eta^2 = 0.04$], as well as a marginally significant interaction of contrast × stimulation × step [$F_{(2,46)} = 3.02$, $p < 0.06$, $\eta^2 = 0.01$].

We further examined the contrast × stimulation interaction in the identity condition separately, for ambiguous (midpoint), and unambiguous (voiceless) steps. The interaction was significant for the unambiguous voiceless steps [$F_{(2,46)} = 3.81$, $p < 0.03$, $\eta^2 = 0.038$], but not for ambiguous ones ($F < 1$).

Finally, we probed the effect of stimulation on response time to unambiguous sounds in the identity condition using planned contrasts. Results showed that, for coronal sounds, the (congruent) tongue stimulation facilitated response time compared to both the incongruent lips stimulation [$t_{(46)} = 2.70$, $p < 0.01$] and control [$t_{(46)} = 3.14$, $p < 0.003$] conditions; stimulation did not reliably affect responses to labial sounds ($t < 1$). Thus, in Experiment 3, congruent stimulation facilitates sensitivity to the contrast between a coronal sound that is ambiguous with respect to its voicing and a voiceless coronal step.

Experiment 4 presented participants with yet a harder discrimination test, as here, each trial featured a subtle contrast between two sounds that were mostly identified as voiceless (steps 2 vs. step 4).

An inspection of the means (**Figure 6**) found lower sensitivity (d') scores relative to previous experiments. But as in those studies, discrimination was once again selectively modulated by stimulation. Sensitivity to the labial contrast decreased when people stimulated the tongue, whereas sensitivity to the coronal contrast decreased when people stimulated the lips.

This impression was supported by a 2 contrast × 3 stimulation ANOVA on d'. The contrast x stimulation interaction was significant [$F_{(2,46)} = 5.12$, $p < 0.01$, $\eta^2 = 0.08$]. No other effects were significant (all $p > 0.4$). Similar ANOVAs conducted on response time found no significant effect for either identity or non-identity trials (all $p > 0.18$, for the means, see **Table 1**).

We next probed the effect of stimulation on the labial and coronal contrasts separately, via planned contrasts. For the labial contrast, the stimulation of the congruent lips articulator significantly improved sensitivity relative to the incongruent

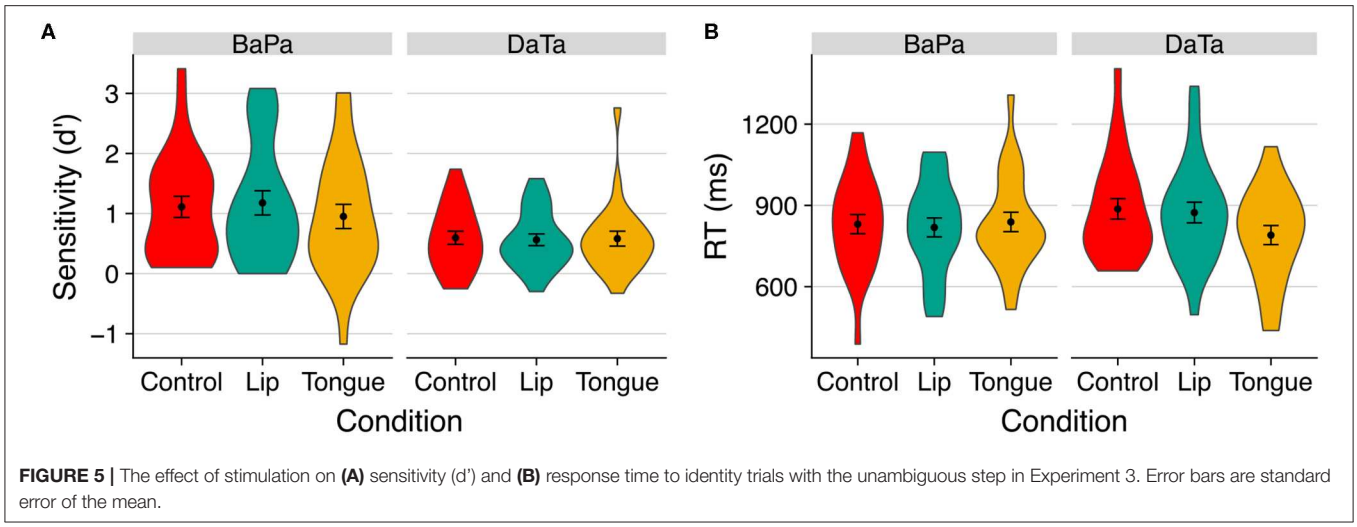

FIGURE 5 | The effect of stimulation on **(A)** sensitivity (d') and **(B)** response time to identity trials with the unambiguous step in Experiment 3. Error bars are standard error of the mean.

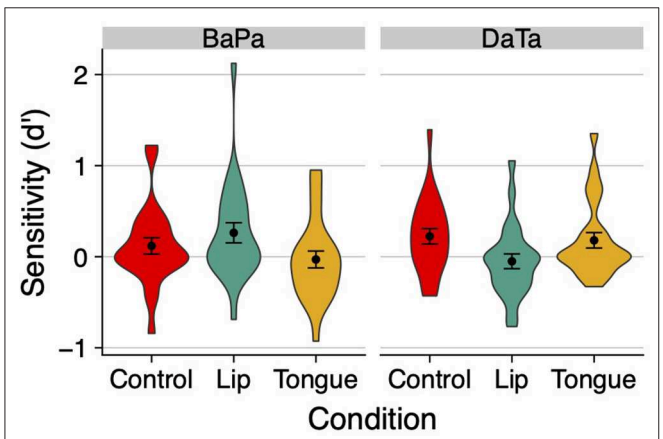

FIGURE 6 | The effect of stimulation on sensitivity (d') in Experiment 4. Error bars are standard error of the mean.

TABLE 1 | The effect of stimulation on response time.

	"Identical" trials			"Non-identical" trials		
	Lips	**Tongue**	**Control**	**Lips**	**Tongue**	**Control**
Mean	768	780	778	906	993	892
SD	197	208	251	229	281	217
Mean	733	778	766	844	842	903
SD	170	251	215	241	279	239

tongue articulator [$t_{(46)} = 2.39$, $p < 0.03$]; responses to the lips and control condition did not differ reliably [$t_{(46)} = 1.22$, $p > 0.23$]. For the coronal contrast, the congruent tongue articulator likewise produced a marginally significant increase in sensitivity relative to the lips [$t_{(46)} = 1.88$, $p < 0.07$], but it did not differ from the control condition ($t < 1$). These results converge with the findings of Experiments 2–3 to suggest that the stimulation of a congruent articulator facilitates the discrimination of a voicing contrast relative to the stimulation of an incongruent articulator.

GENERAL DISCUSSION

A large body of research suggests that the brain's articulatory motor system (e.g., the lip motor area) mediates the identification of speech sounds (for review, Badino et al., 2014; Skipper et al., 2017). But whether this activation is due to motor simulation (e.g., the control of lip movement) or non-motor (e.g., acoustic) computations is largely unexplored.

To elucidate the functional role of the motor system, here we examined the effect of mechanical motor stimulation of the lips and tongue on the perception of sounds that were ambiguous with respect to their voicing. Across experiments, the stimulated articulator (e.g., lips) was always equally relevant (or irrelevant) for the two response options (e.g., *ba/pa*). Each stimulated articulator, however, was congruent with one contrast: the lips were congruent with the labial *ba/pa* contrast, whereas the tongue was congruent with the *da/ta* contrast. Accordingly, our manipulation allowed us to examine the effect of stimulation on the perception of congruent speech sounds independently of its effect on response selection. Our results show that speech perception is systematically modulated by the stimulation of the congruent articulator.

In Experiment 1 (identification), the stimulation of the congruent articulator facilitated the voicing response compared the incongruent articulator. Specifically, for the *da/ta* contrast, people were more likely to select the voiced option when they bit their tongue, whereas for the labial *ba/pa* contrast, they were faster to respond when they bit on their lips.

Experiments 2–4 further showed that, compared to incongruent stimulation, congruent stimulation facilitates the discrimination of a voicing contrast. Experiment 2 found an improved sensitivity at a cost of slower response only for coronal sounds. In Experiment 3, congruent stimulation sped up responses to coronal sounds. Finally, Experiment 4 found

reliable effects of stimulation for both labial and coronal sounds. These were mostly voiceless sounds that subtly contrasted with respect to their voicing. Lip stimulation improved sensitivity to *pa* tokens, whereas tongue stimulation improved sensitivity to tokens of *ta*.

Together, these four experiments show that the stimulation of the congruent articulator modulates speech perception in a selective and systematic fashion. Applying pressure to the lips, for instance, modulates the perception of the labial *ba/pa* contrast relative to the stimulation of the tongue. Moreover, incongruent stimulation reduces perceptual sensitivity, so its effect cannot be due to non-selective strategies of attention, memory, or response selection. Rather, the effects of stimulation are *selectively* modulated by the congruence between the stimulated articulator and the speech sound (e.g., of the lips with *ba/pa*, not *da/ta*). Our results from stop consonants converge with the findings of Masapollo and Guenther (2019) for vowels to suggest that motor stimulation can modulate the perceptual sensitivity of adult listeners[4]. As such, these results suggest that participants might rely on ***motor simulation*** for speech perception.

We note that our conclusions are limited inasmuch as our experimental procedure could have drawn participants' attention to motor articulators, and in so doing, it could have encouraged their reliance on motor simulation. Further research is thus necessary to assess the generality of our conclusions.

How is the process of motor simulation modulated by the mechanical stimulation of the lips and tongue? One possibility is that biting on the congruent articulator improves speech perception by directly stimulating the somatosensory system. In this view, hearing a labial sound requires the engagement of the lip motor system; pressing on the lips stimulates this process, and in so doing, it facilitates the perception of labial sounds. The precise effect of simulation on the perception of voicing is open to two interpretations. One possibility is that the stimulation of the relevant articulator modulates the perception of voicing, perhaps because oral closure and voicing are interdependent (e.g., along the lenis-fortis contrast; for some support, see Colin et al., 2000; Higashikawa et al., 2003)[5]. Alternatively, the perception of a segment could require an alignment of its perceived articulatory gestures; incongruent mechanical stimulation could disrupt this alignment, and in so doing, interfere with perception of all features, including the voicing feature. In all of these plausible scenarios, direct somatosensory stimulation is affecting the perception of voicing, as observed in the experiments reported here.

Another possibility is that mechanical stimulation facilitates motor simulation indirectly, via attention allocation. By focusing attention on the relevant articulator, congruent stimulation might improve motor simulation, whereas incongruent simulation might disrupt it by diverting attention to irrelevant articulators. Whether the effects of mechanical stimulation are due to the stimulation of the motor system directly, to its indirect stimulation via attention allocation, or both is a question that falls beyond the scope of the present research. Our findings, however, clearly show that the effect of mechanical stimulation is *selective*, and as such, they are suggestive of motor simulation.

Furthermore, the convergence between the outcomes of mechanical end-organ stimulation (e.g., of the lips) and brain TMS (e.g., the lip motor area; e.g., Smalle et al., 2014) strongly suggests that their function is shared. These conclusions shed light on the large neuroimaging literature, showing the activation of the brain articulatory motor system in speech perception (e.g., Hickok and Poeppel, 2007; Badino et al., 2014) to suggest that this activation reflects motor simulation, specifically. We thus conclude that phonetic categorization of speech sounds engages articulatory motor action.

Interestingly, our present results from phonetic categorization contrast with our previous investigation of the role of the motor system in phonology. At the phonological level, we found that motor simulation was not necessary for phonological computation, as people remained sensitive to syllable structure even when motor simulation was disrupted, either by TMS (Berent et al., 2015) or by mechanical stimulation (Zhao and Berent, 2018). The dissociation between the positive engagement of the motor system in phonetic categorization (as shown by the present results and past brain findings) and the absence of motor simulation in the computation of phonology (Berent et al., 2015) suggests that the role of the motor system in speech perception might depend on the level of analysis. While phonetic categorization relies on motor simulation, other aspects of phonological computations might be disembodied and abstract.

These conclusions shed light on the ongoing debate concerning the role of embodiment and abstraction in cognition (Barsalou, 2008; Mahon and Caramazza, 2008; Pulvermüller and Fadiga, 2010). The contrasting contributions of motor simulation at distinct levels of speech processing (phonetics vs. phonology) offer a valuable lesson. The finding that motor simulation mediates one level of analysis (e.g., phonetics) does not preclude the role of abstraction at other levels (e.g., phonology). A resolution of the embodiment debate may thus require careful attention to the level of analysis.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

[4]We note, however, response latency in our experiments was longer than in previous research. For example, Pisoni and Tash (1974) report a mean response time of 450–600 ms in an identification task, and roughly 341 ms in a discrimination task, whereas in our experiments, response latency is substantially longer even in the control condition (in Experiment 1: $M = 777$ ms; in Experiments 2-4; $M = 877$ ms). One possibility is that these longer latencies are due to the challenging perceptual conditions. Indeed, our identification task featured a single ambiguous token (not a full continuum, which is mostly unambiguous). Our discrimination experiments were also more demanding, as the ISI was extensive (1,200 ms, compared with 250 ms, in Pisoni and Tash, 1974) and participants were asked to judge physical identity (rather than phonetic identity, in Pisoni and Tash, 1974). Alternatively, responses to the control condition could have been affected by the stimulation conditions, either because the stimulation task encouraged motor simulation throughout the experiment, or because it altered participants' response deadline.

[5]We thank an anonymous reviewer for suggesting this interpretation to us

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, Northeastern University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

IB conceived the study and wrote the initial draft. IB, AP-L, MP, and PF designed the experiments. EB generated the stimuli. MP and RT analyzed the data. All authors offered valuable comments.

FUNDING

This research was supported by NSF grants 1528411(to IB) 1733984 (to IB/AP-L).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2020.00034/full#supplementary-material

REFERENCES

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Badino, L., D'Ausilio, A., Fadiga, L., and Metta, G. (2014). Computational validation of the motor contribution to speech perception. *Top. Cogn. Sci.* 6, 461–475. doi: 10.1111/tops.12095

Barsalou, L. W. (2008). Grounded cognition. *Ann. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., and Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4429–4434. doi: 10.1073/pnas.1014818108

Berent, I., Brem, A.-K., Zhao, X., Seligson, E., Pan, H., Epstein, J., et al. (2015). Role of the motor system in language knowledge. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1983–1988. doi: 10.1073/pnas.1416851112

Berent, I., Zhao, X., Balaban, E., and Galaburda, A. M. (2016). Phonology and phonetics dissociate in dyslexia: evidence from adult English speakers. *Lang. Cogn. Neurosci.* 31, 1178–1192. doi: 10.1080/23273798.2016.1211301

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer*. [computer program] (Version 6.0.37). Available online at: http://www.praat.org/ (accessed August 11, 2019).

Bruderer, A. G., Danielson, D. K., Kandhadai, P., and Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13531–6. doi: 10.1073/pnas.1508631112

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809. doi: 10.1016/j.cognition.2008.04.004

Colin, C., Radeau, M., Demolin, D., and Soquet, A. (2000). "Visual lipreading of voicing for French stop consonants," in *Paper presented at the Proc. Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000*, (Beijing, China).

Cooper, W. E., Blumstein, S. E., and Nigro, G. (1975). Articulatory effects on speech perception: a preliminary report. *J. Phonetics* 3, 87–98.

D'Ausilio, A., Bufalari, I., Salmas, P., and Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* 48, 882–887. doi: 10.1016/j.cortex.2011.05.017

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol. CB* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Elli, G. V., Lane, C., and Bedny, M. (2019). A double dissociation in sensitivity to verb and noun semantics across cortical networks. *Cereb. Cortex.* 29, 4803–4817. doi: 10.1093/cercor/bhz014

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Nuerosci.* 15, 399–402. doi: 10.1046/j.0953-816x.2001.01874.x

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462:502. doi: 10.1038/nature08572

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Higashikawa, M., Green, J. R., Moore, C. A., and Minifie, F. D. (2003). Lip kinematics for /p/ and /b/ production during whispered and voiced speech. *Folia Phoniatr. Logo.* 55, 17–27. doi: 10.1159/000068059

Holt, L. L., and Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Curr. Direct. Psychol. Sci.* 17, 42–46. doi: 10.1111/j.1467-8721.2008.00545.x

Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1245–1248. doi: 10.1073/pnas.0810063106

Kauffman, T., Théoret, H., and Pascual-Leone, A. (2002). Braille character discrimination in blindfolded human subjects. *Neuroreport* 13, 571–574. doi: 10.1097/00001756-200204160-00007

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461.

Linares, D., and López-Moliner, J. (2016). Quickpsy: an R package to fit psychometric functions for multiple groups. *R J.* 8, 122–131. doi: 10.32614/RJ-2016-008

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004

Martino, J., Velasquez, C., Vázquez-Bourgon, J., de Lucas, E. M., and Gomez, E. (2017). Cross-modal recruitment of auditory and orofacial areas during sign language in a deaf subject. *World Neurosurg.* 105, 1033.e1031–1033.e1035. doi: 10.1016/j.wneu.2017.05.170

Masapollo, M., and Guenther, F. H. (2019). Engaging the articulators enhances perception of concordant visible speech movements. *J. Speech Lang. Hear. Res. JSLHR* 62, 3679–3688. doi: 10.1044/2019_JSLHR-S-19-0167

Matchin, W., Groulx, K., and Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *J. Cogn. Neurosci.* 26, 606–620. doi: 10.1162/jocn_a_00515

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Möttonen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi: 10.1523/JNEUROSCI.6018-08.2009

Nasir, S. M., and Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20470–20475. doi: 10.1073/pnas.0907032106

Ogane, R., Schwartz, J.-L., and Ito, T. (2020). Orofacial somatosensory inputs modulate word segmentation in lexical decision. *Cognition* 197:104163. doi: 10.1016/j.cognition.2019.104163

Pant, R., Kanjlia, S., and Bedny, M. (2019). A sensitive period in the neural phenotype of language in blind individuals. *bioRxiv* 592345. doi: 10.1101/592345

Pisoni, D., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.* 15, 285–290. doi: 10.3758/BF03213946

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7865–7870. doi: 10.1073/pnas.0509989103

Rosenblum, L. D., and Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 318–331. doi: 10.1037/0096-1523.22.2.318

Sams, M., Möttönen, R., and Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Res. Cogn. Brain Res.* 23, 429–435. doi: 10.1016/j.cogbrainres.2004.11.006

Sato, M., Grabski, K., Glenberg, A. M., Brisebois, A., Basirat, A., Ménard, L., et al. (2011). Articulatory bias in speech categorization: evidence from use-induced motor plasticity. *Cortex* 47, 1001–1003. doi: 10.1016/j.cortex.2011.03.009

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychol. Sci.* 24, 1824–1830. doi: 10.1177/0956797613478614

Scott, M., Yeung, H. H., Gick, B., and Werker, J. F. (2013). Inner speech captures the perception of external speech. *J. Acoust. Soc. Am.* 133, EL286–92. doi: 10.1121/1.4794932

Skipper, J. I., Devlin, J. T., and 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 Lang.* 164, 77–105. doi: 10.1016/j.bandl.2016.10.004

Smalle, E. H. M., Rogers, J., and Möttönen, R. (2014). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cereb. Cortex* 25, 3690–8. doi: 10.1093/cercor/bhu218

Stokes, R. C., Venezia, J. H., and Hickok, G. (2019). The motor system's [modest] contribution to speech perception. *Psychon. Bull. Rev.* 26, 1354–1366. doi: 10.3758/s13423-019-01580-2

Striem-Amit, E., Cohen, L., Dehaene, S., and Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–652. doi: 10.1016/j.neuron.2012.08.026

Venezia, J. H., Saberi, K., Chubb, C., and Hickock, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Front. Psychol.* 3:157. doi: 10.3389/fpsyg.2012.00157

Wichmann, F., and Hill, N. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313. doi: 10.3758/BF03194544

Zhao, X., and Berent, I. (2018). The basis of the syllable hierarchy: articulatory pressures or universal phonological constraints? *J. Psycholinguist. Res.* 47, 29–64. doi: 10.1007/s10936-017-9510-2

Conflict of Interest: The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, the National Institutes of Health, or the Sidney R. Baer Jr. Foundation. AP-L serves on the scientific advisory boards for Nexstim, Neuronix, Starlab Neuroscience, Neuroelectrics, Axilum Robotics, Magstim Inc., and Neosync; and is listed as an inventor on several issued and pending patents on the real-time integration of transcranial magnetic stimulation with electroencephalography and magnetic resonance imaging.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.