

Postural Control of the Vocal Tract Affects Auditory Speech Perception

H. Henny Yeung

Simon Fraser University and CNRS, University of Paris

Mark Scott

Qatar University

Many researchers have proposed that sensorimotor information about the dynamic production of speech gestures can supplement the auditory perception of speech. Here we show that information about postural, nonspeech control of the vocal tract—such as breathing through the nose or mouth—also affects speech perception. Experimental participants breathed either through the nose or the mouth while identifying categories of speech sounds differing in nasal versus oral airflow. Participants showed an increased tendency to hear speech sounds as having nasal articulation when breathing through the nose, relative to when breathing through the mouth. These results suggest that postural information about the state of the vocal tract, like the motor configuration of the speech articulators while breathing, can modulate the perceptual processing of speech sounds.

Keywords: speech, perception, production, vocal tract, breathing

The cognitive processes involved in perceiving and producing speech are closely intertwined. For example, successful speech production relies on auditory feedback, as shown by the deteriorating quality of spoken speech after hearing loss (Cowie & Douglas-Cowie, 1992). Experimentally manipulating auditory feedback while speaking can similarly affect the execution of vocal gestures (Houde & Jordan, 1998, 2002; Lee, 1950; Nasir & Ostry, 2008; Tourville, Reilly, & Guenther, 2008; Yuen, Davis, Brysbaert, & Rastle, 2010). Overall, there is wide consensus that acoustic-phonetic feedback plays a vital role in a variety of speech production processes including speech motor control (Guenther & Vladusich, 2012; Perkell et al., 2000), speech planning (Levelt, 2001), and speech error correction (Dell & Chang, 2014; Nozari, Dell, & Schwartz, 2011).


Speech researchers sometimes argue, complementarily, that perception draws upon speech production information. For example, early proponents of “motor theories” argued that auditory speech was decoded and understood with gestural/articulatory information

(Galantucci, Fowler, & Turvey, 2006; Liberman, 1996; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Recent theoretical positions have been more cautious, however, as speech-motor deficits do not consistently trigger a loss in the quality of speech perception (Bishop, Brown, & Robson, 1990; Hickok, Costanzo, Capasso, & Miceli, 2011), and conflicting experimental findings have generated significant debate within the literatures on neuroscience (Hickok, Houde, & Rong, 2011; Pulvermüller & Fadiga, 2010; S. K. Scott, McGettigan, & Eisner, 2009; Skipper, Devlin, & Lametti, 2017) and cognitive science (Diehl, Lotto, & Holt, 2004; Hickok, 2014; Schwartz, Basirat, Ménard, & Sato, 2012) as to whether, and to what degree, ordinary speech perception draws upon gestural processes.

An emerging view in the perception literature is that motor influences play a supplementary role in speech perception. For example, Pickering and Garrod (2007) argue that top-down predictions derived from the motor system can “fill in the gaps” of the auditory signal, constraining the possibilities entertained in perception, thus easing processing load. Skipper and colleagues (Skipper, Goldin-Meadow, Nusbaum, & Small, 2007; Skipper, Nusbaum, & Small, 2006) have proposed a similar theory, arguing that the motor system is activated in difficult listening situations, which could model auditory input and thus aid speech perception. Schwartz et al. (2012) also argue for this supplementary role of motor information, but additionally suggest that the motor system shapes speech acquisition, with phonetic learning constrained by information about speech production. Hickok and colleagues have similarly proposed a framework for possible neural architectures that might instantiate these motor-influenced pathways in speech perception (Hickok, Houde, et al., 2011; Hickok & Poeppel, 2007).

The current study further explores sensorimotor modulation of speech perception, asking what format motor information can take as it influences perception. We draw a distinction between two types of motor information: First, postural information about speech articulation, and second, the execution of dynamic speech gestures. We refer to postural information as the active mainte-

This article was published Online First October 29, 2020.

 H. Henny Yeung, Department of Linguistics, Simon Fraser University, and Integrative Neuroscience and Cognition Center (UMR 8002), CNRS, University of Paris; Mark Scott, Department of English Literature and Linguistics, Qatar University.

Portions of the present work, including early analyses of portions of this data as well as brief write-ups of these ideas, have been presented at two prior conferences: the 10th International Conference on Cognitive Science in 2015, and at Acoustic Week in Canada in 2016. This research was funded in part by grants to H. Henny Yeung from the LABEX-EFL in France (ANR-10-LABX-0083) and NSERC in Canada (RGPIN-2018-04990). The authors thank Sylvie Margules and Elena Berdasco for assistance in data collection, as well as Lionel Granjon for comments and advice on analysis.

Correspondence concerning this article should be addressed to H. Henny Yeung, Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada. E-mail: henny_yeung@sfu.ca

nance of vocal tract articulators in certain positions, which is distinct both from vocal tract rest positions (Gick, Wilson, Koch, & Cook, 2004; Ramanarayanan, Lammert, Goldstein, & Narayanan, 2014), and from the dynamic processes involved in speech segment articulation (Tilsen et al., 2016). While most prior work investigating motor influences on perception have explored the latter (i.e., dynamic processes), our results are novel in showing that motor information about vocal tract posture, or the simple maintenance of a particular vocal tract configuration—that is, whether one is breathing through the mouth or through the nose—can be incorporated into judgments about speech perception. These results change our understanding of what kinds of gestural information are integrated into perceptual processes, expanding notions of what kinds of motor information are used in speech production-perception interactions. Broadly, these results lend support to a view of these interactions as being ubiquitous and automatic, and spanning multiple levels of the motor hierarchy.

The Role of Speech Gestures in the Perception of Speech

As reviewed above, several theories of speech perception support the claim that some mechanisms of perception rely upon gestural information to analyze the speech signal, particularly when speech is ambiguous or hard to hear (Hickok, Houde, et al., 2011; Schwartz et al., 2012). Experimental studies bolster this idea by showing that speech gesture information influences perception under ambiguous listening conditions. For example, applying transcranial magnetic stimulation (TMS) to brain areas controlling particular vocal tract articulators—like the lips or tongue—can bias behavioral judgments about ambiguous speech sounds (D'Ausilio et al., 2009; Möttönen, Dutton, & Watkins, 2013; Möttönen & Watkins, 2009, 2012; Sato, Tremblay, & Gracco, 2009). Other work has similarly shown that silently mouthing dynamic articulatory gestures, whether speech-like (Mochida et al., 2013; Sams, Möttönen, & Sihvonen, 2005; M. Scott, 2013; M. Scott, Yeung, Gick, & Werker, 2013) or non-speech-like (Sato et al., 2011), can have a similar effect on auditory speech perception. Yet another set of studies has investigated adaptation and learning, showing that inducing temporary changes in the way that speech gestures are executed will result in concomitant modulation of speech perception (Lametti, Krol, Shiller, & Ostry, 2014; Shiller, Sato, Gracco, & Baum, 2009).

Contemporary explanations for these effects have largely appealed to the concept of *forward models*, which are frequently used in cognitive science to explain the integration of sensory and motor information. Forward models are a component of the motor system that generates predictions about the upcoming sensory consequences of a planned motor action. These predictions are then compared with actual sensory feedback from the executed action for the purposes of error correction and future motor planning (Flege, Takagi, & Mann, 1996; Kawato, 1999; Wolpert, Ghahramani, & Jordan, 1995). Because forward models predict the sensory consequences of an action, while also taking into account sensory feedback from the executed action, a common hypothesis in this literature is that perceivers will perceive ambiguous sensations as percepts in line with expectations generated from a forward model (Schütz-Bosbach & Prinz, 2007). For example, Repp and Knoblich (2007) demonstrated a kind of motor-induced “per-

ceptual capture” by having pianists perform hand motions along with an ambiguous sequence of notes that could be heard either as rising or falling. When performing the hand motion consistent with a rising sequence, pianists tended to hear the ambiguous sound sequence as ascending (and vice versa).

In the speech domain, researchers similarly suggest that forward models predict the sensory consequences of a planned vocal tract gesture, which are then evaluated against sensory feedback about the executed gesture (Greenlee et al., 2011; Hickok, Houde, et al., 2011; Schwartz et al., 2012). Behavioral studies provide support for both the predictive and feedback-related aspects of these models. For example, speech planning—simply intending to produce certain speech gestures—generates a sensory prediction that can modulate speech perception (Berger & Ehrsson, 2013; M. Scott, 2013; M. Scott et al., 2013). Likewise, incoming feedback from somatosensory receptors in skin and muscle—such as when a machine provides artificial tactile input to the subject about facial skin deformation when one is speaking—similarly triggers a perceptual modulation of speech (Ito, Tiede, & Ostry, 2009).

Theoretical models that describe the neurobiological basis of speech perception also appeal to forward models to explain why neural activation in speech perception tasks are often seen in both “auditory” and “motor” areas of the brain. For example, activations of dorsal-posterior networks and portions of inferior frontal gyrus (Broca’s area) when processing speech information are commonly thought to reflect the functioning of a forward model that offers a categorical speech structure with which to compare graded auditory activation (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Rauschecker & Scott, 2009). Others interpret this type of activation as a supplement to speech perception in cases of degraded auditory activation (Hickok, Houde, et al., 2011; Skipper et al., 2007, 2006). Together with behavioral evidence, these studies lend support to accounts of speech processing that link perception with speech production through a framework using forward models.

Defining Gestural Information in Forward Models of Speech Processing

While significant progress has been made in describing pathways linking speech perception and production through forward models, relatively little is known about the format of sensorimotor information. Current debates about this issue have centered on the question of how “speech-like” sensorimotor information must be in order to be considered perceptually relevant by forward models.

One position is that sensorimotor information must match the temporal profile of a speech gesture in order to be successfully integrated into perceptual processing. For example, in a study exploring afferent¹ (incoming) sensory information about speech production—when an external source deformed the facial skin of a perceiver in ways similar to speaking—perceptual judgments about speech were modulated only when the afferent cues matched the temporal dynamics of a real speech gesture, but not when these cues were out of synchrony (Ito et al., 2009). Other investigations

¹ *Afference* is an incoming sensory signal from any source, but when this sensory signal carries sensory feedback about one’s own body’s position or actions, the correct term is *reafference*. Later in this article we use *reafference* when we refer to feedback processes for incoming sensory signals.

have similarly explored afferent somatosensory cues about speech production, and results also reported perceptual effects that vary with the precise timing between somatosensory information and speech stimuli (Gick & Derrick, 2009; Gick, Ikegami, & Derrick, 2010). Overall, then, proponents of this position could suggest that afferent information from somatosensory cues will require that gestural information be formatted in ways that reflect the temporal, dynamic structure of actual speech gestures.

Similar constraints do not appear to operate on efferent (outgoing) aspects of forward models. For instance, prior work suggests that efferent sensorimotor information does not have to be speech-specific (D'Ausilio et al., 2009), or even time-locked to an auditory stimulus (Möttönen et al., 2013; Möttönen & Watkins, 2009) to exert effects on perception. These questions are typically tested in TMS protocols, which involve stimulation of a part of cortex linked to motor control of a specific vocal tract articulator (e.g., the tongue). This methodology thus isolates efferent activation from reafferent information, since perceivers do not typically move their articulators in these studies, and thus experience little to no somatosensory feedback (Möttönen & Watkins, 2012). In sum, contrary to the somatosensory literature discussed in the prior paragraph, TMS evidence suggests instead that efferent gestural information does not have to be speech-specific, and may not have to be formatted with the same temporal profile as a real speech gesture.

Another, third type of relevant study requires that experimental participants make their own motor gestures, and so activate both efferent channels associated with a motor command, and reafferent channels associated with perceptual feedback from that motor command. For example, Sato and colleagues (2011) asked experimental participants to execute a series of nonspeech gestures in a training phase, making repeated “kissing” or “tongue-raising” gestures. After the training phase in Sato et al. (2011), participants showed a response bias in perception for identifying either auditory /p/ or /t/, depending on what kind of gesture they had made.

Together, these three types of studies present very different views of what sorts of gestural information are integrated in speech processing, and therefore what the format of motor information in forward models may be. Broadly speaking, there is no consistent consensus on what kinds of motor information count: Studies focusing on reafferent sensory feedback from somatosensory/tactile sources suggest that sensorimotor cues must be dynamically specified in speech-like ways, while TMS studies using efferent information hint at the involvement of broad-based motor information in speech processing that is not speech-like in its temporal dynamics. Studies that involve the executing of articulatory gestures (i.e., creating both efferent and reafferent activation) also suggest that motor information in forward models are specified in ways abstract enough to include both nonspeech and speech gestures.

Overall, the field is left without a clear understanding of what limits or constraints there are with respect to the format of sensory and motor influences on speech perception. Moreover, most current theoretical descriptions do not commit to any specific format with respect to sensorimotor information, relying on a default assumption that phonemic or syllabic information is a likely representational format for perception–production interactions, with lower-level sensorimotor information taking a less prominent role (Hickok, Houde, et al., 2011; Tourville & Guenther, 2011). De-

fining the precise format of sensorimotor information in forward models will be a challenging task for future research, but in this single study we explore an important boundary condition. Here we ask whether postural control about the static position of the articulators (in both efferent and reafferent channels) is sufficient to be integrated into speech perception.

Overview of the Current Study

A requirement of forward models is that all relevant information about the positions and movements of vocal tract articulators are accounted for, which, we argue, even includes postural information about static articulator positions. This idea is based on the assumption that sensory predictions must take into account the starting point of an action in order to be accurate (Hickok, 2012b; Hickok, Houde, et al., 2011). For example, moving one's finger toward the body will generate different sensory predictions when the starting point is at chest-level (just touching the finger to the chest), as opposed to eye-level (a painful finger touch to the eye).

A prediction from this idea is that forward models in the speech domain are sensitive to the actively maintained positions of the speech articulators, even if one is not executing the dynamic gestures required to produce auditory sounds when speaking (Gick et al., 2004; Ramanarayanan et al., 2014; Tilsen et al., 2016). Thus when a person's velum is down in a gesture that permits nose breathing, predictions from a forward model should take this information into account, which could thus bias a perceiver toward perceiving nasality (sounds produced with a lowered velum) when hearing ambiguous speech. In the present study, we test this hypothesis by asking whether speech perception is modulated simply by whether one is instructed to breathe through the nose or through the mouth.

Nose and mouth breathing gestures were tested for several reasons: First, we almost always breathe through the mouth and/or nose when listening to speech, and so our study explores an everyday, ecological occurrence. Second, the neural sources of control for speech and for respiration overlap in some crucial ways, as one must control airflow through the vocal tract when speaking (Ackermann & Ziegler, 2010; Jürgens, 2002). Our data may thus help to link our understanding of the basic neural processes underlying speech on the one hand, and respiration on the other. Finally, mouth and nose breathing differ principally in the position of a single superlaryngeal articulator: the velum, a flap of tissue at the back of the mouth, which helps direct airflow through the nasal or oral cavities (Figure 1 shows an approximation of this), while also being the principal articulator in distinguishing nasal and oral consonants (Bell-Berti, 1993).

In this study, participants categorized auditory sounds from a speech continuum while maintaining either nasal or oral respiratory configurations. There were two conditions in this study, the DN and DG conditions, which differed only in the type of auditory continuum used.

In the *DN condition*, the continuum endpoints were disyllables containing consonants articulated with either a raised velum (/ada/) or lowered velum (/ana/). The articulatory targets for these consonants are globally similar to the two target breathing gestures in terms of articulator configuration: The articulation of the oral consonant /d/ is similar to mouth breathing in having a raised velum with contact between the velum and the pharyngeal wall; by

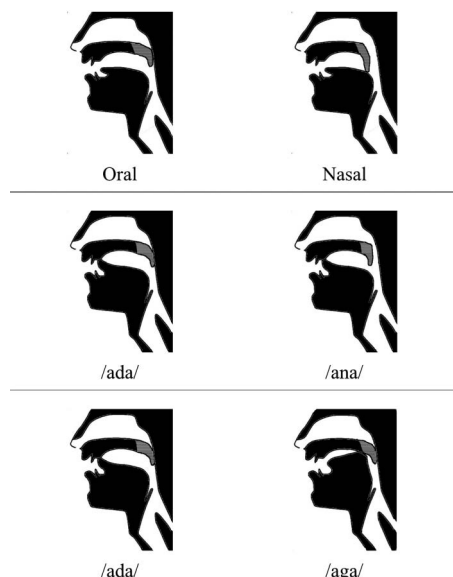


Figure 1. Vocal tract configurations highlight the position of the velum (gray shading) during oral and nasal respiration with an open mouth (top row). Here, the velum either touches the pharyngeal wall or the back of the tongue in order to direct airflow through oral or nasal cavities. Sounds tested in the DN condition (middle row) consisted of an auditory /ada/-/ana/ contrast, which involves articulations that also differ in velar-pharyngeal contact and in airflow through the vocal tract. In the DG condition (bottom row), an auditory /ada/-/aga/ contrast was tested, which involves articulations that do not capture the differences in oral and nasal breathing as much, with respect to velum position and airflow through the vocal tract.

contrast, the articulation of /n/ is similar to nose breathing in having a lowered velum without velar-pharyngeal contact, but with velar-tongue contact (Perry, 2011). The aerodynamics of oral and nasal consonants, like /d/ and /n/, are more complex, as airflow is present in both nasal and oral cavities: Nasalization simply alters the relative impedance of airflow between these two regions of the vocal tract (House & Stevens, 1956; Warren, Dalston, & Mayo, 1993). There are nevertheless gross similarities between consonant articulation and breathing. Producing /d/ results in greater impedance of nasal airflow, approximating mouth breathing, while producing /n/ more significantly impedes oral airflow, approximating nose breathing.

In the *DG condition*, the endpoints of the target continuum were /ada/ and /aga/. As seen in Figure 1, consonant articulations in this condition did not align well with the gestural configurations of breathing gestures. In terms of articulator configurations, both /d/ and /g/ involve a raised velum with contact between the velum and the pharyngeal wall (recall that /n/ does not typically involve velar-pharyngeal contact). In terms of airflow, both /d/ and /g/ articulations involve relatively greater impedance of nasal airflow than /n/.

In sum, these experimental conditions explored the idea that breathing in a particular way affects the perception of consonants. As a preview of the results, Experiments 1 and 2 (two versions of the same basic procedure) showed that breathing through either the nose or mouth modulates speech processing, but only in the DN

condition, where the experimental task was to distinguish the articulation of oral and nasal consonants. Together, these experiments show that sensorimotor influences on speech perception are automatic, and can involve information about postural control over the vocal tract articulators, even ones that are not actively being used to produce speech. This lends support to an emerging view of broad and ubiquitous interactions between motor and perceptual information in the speech domain (Skipper et al., 2017). As discussed in the final section of this article, these findings have broad implications for theories of speech perception and production, as well as our understanding of clinical speech disorders.

Method

In both experiments, native speakers of French identified auditory speech sounds from one of two 10,000-step speech continua generated from naturally produced French disyllables using speech synthesis software (STRAIGHT; Kawahara, Irino, & Morise, 2011). Speech tokens came either from an /ada/-/ana/ continuum (the DN condition), or from an /ada/-/aga/ continuum (the DG condition) in alternating experimental blocks. Crucially, participants also controlled their breathing during the categorization task, breathing through their nose in some blocks, and through their mouth in others.

We asked whether the perceptual “boundary” between the two speech categories was affected by participants’ respiratory gestures, and these boundaries were assessed using an adaptive staircase procedure (Amitay, Irwin, Hawkey, Cowan, & Moore, 2006; Cornsweet, 1962; Leek, 2001). Two simultaneous staircases were run within each experimental block of trials, and each staircase procedure began with the presentation of stimuli close to one end of the continuum (e.g., /ada/ for one staircase; /ana/ for the other). Subsequent stimuli presentations were progressively closer to the other endpoint, and presented with a particular “step size” along the continuum. Once category responses changed, stimuli were then presented in smaller steps back toward the original end of the continuum. This repeated until the step size reached a predetermined minimum value. The perceptual boundary was then operationalized as the average point along the continuum where the last six “reversals” occurred on each staircase, following previous work (Amitay et al., 2006; Cornsweet, 1962; Leek, 2001).

Participants

Eighty-eight native monolingual French speakers were recruited from the community at the University of Paris, and were given 10€ compensation for participation: Experiment 1 had 40 participants (10 male; 30 female; mean age = 22.4 years); Experiment 2 had 48 participants (11 male; 37 female; mean age = 22.5 years). While a sample size of 40 was planned for both studies, all those who responded to e-mail study advertisements were tested, resulting in sample sizes over this number, which was then reduced to the largest sample size (in Experiment 2) that would still respect the counterbalancing of experimental orders. Fourteen additional participants were tested, but data was either recorded incorrectly due to computer programming errors ($n = 12$), or not analyzed because their inclusion would have disrupted the counterbalancing order ($n = 2$). The ethics committee at the University of Paris approved this research.

Stimuli

The continuum in the DN condition was made by resynthesizing two naturally produced /ada/ and /ana/ speech tokens, which were recorded from a female native speaker of French, and interpolating 10,000 steps between them. In the DG condition, a synthesized continuum was also made in the same way, based on naturally produced /ada/ and /aga/ tokens from the same speaker. All continua were generated using STRAIGHT, an auditory synthesis and morphing program (Kawahara et al., 2011). The two 10,000-step continua were generated using a log-linear scale between the spectra of the endpoints. Thus, there is no blending (i.e., STRAIGHT does not use a weighted overlap between endpoints), but instead, all stimuli consisted of a calculated interpolation between spectral endpoints. The amplitudes of the endpoints were normalized before continua generation, using the parameters available in STRAIGHT, and all steps were set to have equal pitch, intensity, and temporal profile. Moreover, after stimuli were generated, scripts in Praat (Boersma, 2001) were used to renormalize all tokens to the same intensity, and leading and trailing silences were edited to start and end at zero-crossings in the waveform.

General Procedure

Participants were told that the study was investigating the effects of mindfulness on perception, and that they would be asked to control their breathing while performing a simple auditory task. All were debriefed at the end of the study about the actual research question, which no participant correctly guessed. The procedure lasted about 20–30 min.

The main perception task involved a series of experimental trials, which involved the presentation of a sound from one of the stimuli continua. Participants had to identify this sound by pressing one of two buttons on a computer keyboard, with the subsequent trial beginning 150 ms after the button press. At the beginning of the tasks, participants could adjust the volume to their comfort level in the first few experimental trials, while an experimenter monitored responses to ensure that the participant followed prompts, but sound presentation was constrained to a particular intensity range (about 60–65 dB) and fixed for the remainder of the task.

Breathing through the nose or mouth involves forcing air through vocal tract cavities of different sizes, and so one concern is that breathing in a certain way could create differential levels of turbulent, fricative noise. We considered the potential role of auditory interference from breathing, but three reasons mitigated worries about this factor. First, sounds were played at a comfortable volume for each participant through headphones, but constrained to a particular intensity range (about 60–65 dB): Because breathing noise is typically more than 30 dB quieter (Seren, 2006), breathing noise would be substantially masked by the presentation of auditory stimuli. Second, the higher frequencies of turbulent airflow are not concentrated in the same range as the spectral cues that distinguish nasal from oral resonance (Chen, 1997). Most crucially, there should be no systematic similarity between the fricative-like noise from either nose or mouth breathing, and the acoustic cues that distinguish nasal from oral consonants. For nose breathing, turbulence occurs at the nostrils as airflow is forced through a narrow opening. This noise generation occurs outside the nasal cavity, and so there is little filtering of that noise by that

cavity. For mouth breathing, turbulent air is filtered by the oral cavity, but the shape of this cavity when the mouth is open to breathe is very different from its shape when articulating the oral consonants tested here (/ada/ and /aga/), precluding systematic similarity between this breathing sound and the spectrum of either speech sound. Together, all of these reasons minimized concern that turbulent breathing noise might influence perceptual judgments.

Experiment 1. The main experimental task consisted of eight blocks of trials, and at the beginning of each block participants were told whether they should breathe through their mouth or nose and what sounds they would be identifying. When breathing through their nose, participants were asked to keep their mouths open, which ensured that participants' vocal tract configurations differed principally in the position of the velum. Participants were also shown that an experimenter would be observing them over a closed-circuit video camera in order to encourage attention to the experimental task.

Experimental trials were organized into four types of trial blocks: DN condition-mouth breathing; DN condition-nose breathing; DG condition-mouth breathing; and DG condition-nose breathing. One of each trial block was presented, which was then followed by a break of a few minutes, and then each trial block was tested again in the same order (giving a total of 8 trial blocks). The order of presentation of each trial block was counterbalanced across participants in either a nose-first order (nose-nose-mouth-mouth) or mouth-first order (mouth-mouth-nose-nose). The order of DN or DG conditions was also counterbalanced across participants within each pair of nose or mouth trial blocks. In addition, the assignment of the /ada/ speech category to a response key (either the left or right arrow on the keyboard) was constant across all blocks within each participant, but counterbalanced across participants. This created a total of eight unique experimental orders, and as mentioned above, a sample size of 40 was planned with five participants in each order.

Within each block, participants completed a staircase procedure (Cornsweet, 1962) on either an /ada/-/ana/ or /ada/-/aga/ continuum (in the DN and DG conditions, respectively). The staircase consisted of two interleaved staircases, randomly presented trial to trial. If one staircase completed before the other, remaining trials tested the incomplete staircase. The "low" staircase began at Token 2,400 on the continuum, and a "high" staircase began at Token 7,600 on the continuum. In each staircase, there were 13 step sizes: 1,250, 1,000, 800, 650, 500, 350, 250, 150, 80, 50, 30, 20, and 10. Thus, trials related to the lower staircase might begin with the presentation of Token 2,400, then Token 3,550, and so forth until the response switched from /ada/ to /ana/, at which point the staircase would reverse, using step size 1,000 to move back down the continuum. This would continue until 13 reversals were observed, at which point the staircase procedure would be complete. In each trial block, two perceptual boundaries were estimated from the last six reversals of the high and low staircases. After a short break of a few seconds, participants would begin a new block of trials with new task instructions, until all eight blocks were completed.

Experiment 2. The procedure here was identical to Experiment 1, except for a few minor differences. First, the interleaved staircase procedure used 12 step sizes: 1,600, 800, 500, 300, 200, 100, 50, 25, 12, 6, 3, and 1. This staircase procedure narrowed

more quickly into the most ambiguous region of the continuum (12 steps instead of 13), and permitted greater precision in measuring the perceptual boundary by allowing smaller step sizes on the continuum. Second, all participants who responded to an advertisement were run, resulting in a slightly larger sample size compared to Experiment 1 ($n = 48$). Third, we also examined the effect of keeping the mouth open or closed while nose breathing, although the written instruction to breathe through the nose remained the same. Participants were instructed to keep their mouth closed ($n = 24$), or instructed to keep their mouth open, like participants in Experiment 1 ($n = 24$). An experimenter also monitored participants' compliance with these instructions over video surveillance.

In a preliminary analysis, we looked for an effect of mouth opening using a linear mixed-effects model that predicted estimated continuum boundaries, which was specified as closely as possible to the model for the main analysis. Specifically, we examined the interaction of the crucial within-subjects factors of Breathing (mouth or nose) and Continuum (DG or DN) with the between-subjects effects of Mouth Position (open or closed during nose breathing) and Block Order (whether nose blocks occurred before or after mouth blocks). In addition, within-subjects main effects of Experiment Half (1st or 2nd) and Staircase (low or high) were included to provide the model multiple estimated boundaries for each of the critical comparisons. All fixed effects were contrast coded (i.e., as 0 or 1), and the random effects structure was maximally specified with participant as a random intercept paired with all possible random slopes. This random effects structure was then reduced if model convergence was an issue in the same way as for the main analysis (see below), and then repeated until the model converged (Barr, Levy, Scheepers, & Tily, 2013). In this case, the most complex model that would converge had just random slopes for Continuum and for Experiment Half. Satterthwaite estimates of degrees of freedom were used to estimate significance using the summary function from the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017), and critically, the theoretically predicted interaction between Breathing and Continuum did not vary as a function of Mouth Position ($p = .52$). Due to this preliminary result, and because we did not manipulate mouth position in Experiment 1, we ignored this factor in the main analysis. Stimuli files and raw data are available to download at <https://osf.io/jqrf3/>.

Results

The primary effect of interest was the influence of breathing—either through the nose or mouth—on the estimated perceptual boundaries from two speech continua (the DN or DG continua). To investigate our theoretically predicted interaction between the factors of Breathing and Continuum, we used a linear mixed-effect model to predict the estimated boundaries from each staircase procedure in each trial block, and included the within-subjects main effects of Breathing and Continuum as well as their interaction. The model also included three-way interactions between those two critical effects and other between-subjects factors: Experiment (1 or 2) and Block Order (whether nose blocks occurred before or after mouth blocks). Finally, we also included the within-subjects main effects of Experiment Half (1st or 2nd) and Staircase (low or high) to help the model account for the multiple boundaries

estimated from all trial blocks. In other words, the model could account for differences between high and low staircases, as well as for differences in trials from the first and second halves of the experimental procedure (which repeated the presentation of the same trial blocks).

All these fixed effects were contrast coded, with the random effects structure maximally specified, using participants as a random intercept (Barr et al., 2013). We then reduced the random effects structure in the following way if the more complex model did not converge: We first removed random slopes for higher-order interactions, then decorrelated random slopes and intercepts, and finally removed the slope that resulted in the smallest reduction in the Akaike information criterion/Bayesian information criterion (AIC/BIC). We repeated this sequence of steps until the most complex model that would converge included just random slopes for the main effects of Breathing, Continuum, and Experiment Half. The structure of the final model is outlined in Table 1. Results are also reported in that table for each fixed effect and interaction, which were estimated with Satterthwaite approximations for degrees of freedom using the summary function from the lmerTest package in R (Kuznetsova et al., 2017).

Model results show that the critical interaction between Breathing and Continuum was significant, $\beta = -418.63$, $p = .006$, but also that there was a significant three-way interaction involving Breathing, Continuum, and Block Order, $\beta = 722.36$, $p < .001$, and a significant two-way interaction involving Breathing and Block Order, $\beta = -1,267.97$, $p < .001$. Figure 2 illustrates these complex interactions. All post hoc tests that analyzed these interactions were Bonferroni-corrected and were conducted using the estimated marginal means from the main model (calculated using the emmeans package; Lenth, 2020). In those post hoc tests, we used Satterthwaite estimates for degree of freedom as done for the main model, again from the lmerTest package (Kuznetsova et al., 2017).

When averaging across both continua, perceptual boundaries were shifted toward /ada/ (i.e., participants made fewer /ada/ judgments in the ambiguous part of the continua) in later blocks compared to earlier ones, which was captured in the Breathing and Block Order interaction. That is, when nose-breathing blocks occurred after mouth-breathing ones, the perceptual boundaries in nose-breathing blocks were shifted toward /ada/ by an average of 574 ($SE = 73.1$) continuum steps, $t(88.2) = 4.56$, $p < .001$. Likewise, when mouth-breathing blocks occurred after nose-breathing ones, perceptual boundaries in mouth-breathing blocks were shifted toward /ada/ by an average of 333 ($SE = 73.1$) continuum steps, $t(88.2) = 7.85$, $p < .001$. Together, these results show that repeated presentation of /ada/ across both continua resulted in a narrowing of the /d/ perceptual category in later blocks. This reflects the well-understood (and statistically large) effect of selective speech adaptation, which is when repeated presentation of a speech sound results in a narrowing of perceptual boundaries around that sound (Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973).

As the three-way interaction between Breathing, Block Order, and Continuum shows, however, this selective speech adaptation effect (again, captured as the above-described interaction between Breathing and Block Order) was only significant when averaging across both continua, and was differentially observed in each continuum (see Figure 2). On the DG continuum, selective speech

Table 1
Results for Fixed Effects From a Model Estimating Perceptual Boundaries

Fixed effect	Estimate	SE	df	t-value	p
Intercept	3453.97	254.88	94.72	13.55	<.001**
First order effects					
Breathing	537.04	119.23	237.45	4.50	<.001**
Continuum	-219.75	378.57	95.50	-0.58	.56
Experiment	493.79	283.61	92.07	1.74	.085
Block Order	598.84	282.43	92.07	2.12	.037*
Experiment Half	-827.70	58.61	88.00	-14.12	<.001**
Staircase	89.69	42.25	1056.02	2.12	.034*
Second order effects					
Breathing × Continuum	-418.63	151.15	1056.02	-2.77	<.001**
Breathing × Experiment	-19.26	133.80	237.25	-0.14	.89
Continuum × Experiment	-552.74	424.94	95.46	-1.30	.20
Breathing × Block Order	-1267.97	133.25	237.25	-9.52	<.001**
Continuum × Block Order	-477.47	423.18	95.46	-1.13	.26
Higher order effects					
Breathing × Continuum × Experiment	59.72	169.69	1056.02	0.35	.72
Breathing × Continuum × Block Order	722.36	168.99	1056.02	4.28	<.001**

Note. Participant $n = 88$. The syntax for the final model, as run using the lme4 package in R, was: Boundary \sim Breathing \times Continuum \times Experiment + Breathing \times Continuum \times BlockOrder + ExperimentHalf + Staircase + (1 + Breathing + Continuum + ExperimentHalf | Participant).

* $p < .05$. ** $p < .001$.

adaptation was observed, such that the perceptual boundary moved toward the /d/ category in later blocks. When looking just at participants who were in the mouth-first block order ($n = 44$), the perceptual boundary was shifted toward the /d/ category in nose-

compared to mouth-breathing conditions by 741 steps ($SE = 94.5$), $t(237) = 7.84$, $p < .001$. Likewise, for participants in the nose-first block order ($n = 44$), the perceptual boundary moved toward the /d/ category by 527 steps ($SE = 94.5$) in the mouth- compared to nose-breathing conditions, $t(237) = -5.58$, $p < .001$.

However, for the DN continuum, this speech adaptation effect—again, a bias for the perceptual boundary to shift toward the /d/ category in later blocks—was asymmetrical. This pattern is interpreted as the interaction of the selective speech adaptation effect with an effect of breathing, such that the perceptual boundary shifts toward the /d/ category (i.e., a tendency to hear /ada/ less often) when breathing through the nose versus the mouth. Thus, when selective speech adaptation and this breathing effect worked together (i.e., when the nose-breathing blocks occurred after mouth-breathing blocks; $n = 44$), there was a significant shift of 407 continuum steps ($SE = 94.5$) toward the /d/ category when breathing through the nose compared to the mouth, $t(237) = 4.31$, $p < .001$. In contrast, when the effect of selective speech adaptation worked against the breathing effect (i.e., when the nose-breathing blocks occurred before the mouth-breathing blocks; $n = 44$), no difference in perceptual boundaries across breathing conditions was observed in this case of conflict, $t(237) = -1.47$, $p = .14$.

This breathing effect can be seen again when collapsing over the (counterbalanced) effect of Block Order. As shown in Figure 3, we observed a significant interaction between Breathing and Continuum, and a post hoc analysis of this effect revealed again that the perceptual boundaries in the nose-breathing condition were smaller than in the mouth-breathing condition (i.e., the difference between DN boundaries in mouth- and nose-breathing conditions was positive). This suggests that perceptual boundaries moved toward the /d/ category an average of 134 ($SE = 67$) steps on the DN continuum relative to the mouth-breathing condition, $t(236) = 2.00$, $p = .046$, which means that participants heard fewer /ada/ sounds when nose-breathing. In contrast, there was no significant shift for breathing on the DG continuum, $t(236) = 1.60$, $p = .11$.

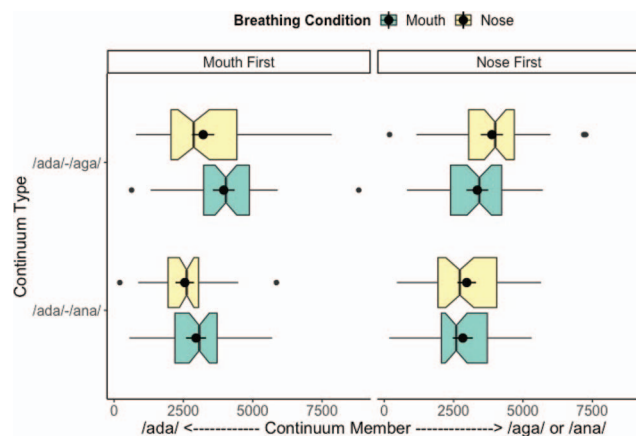


Figure 2. Results here illustrate of the interaction between Breathing (through the mouth or nose), Continuum (/ada/-/aga/ or /ada/-/ana/), and Block Order (whether mouth- or nose-breathing blocks were first) on estimated perceptual boundaries. Box plots show the distribution of data across all subjects, and the notches indicate a distributional estimate of the 95% confidence intervals (CIs; i.e., $\pm 1.58 \times$ the interquartile range divided by the square root of n). The larger dark circles indicate estimated marginal means from the statistical model (with 95% CIs as error bars on those circles). For the /ada/-/aga/ continuum, there was an overall selective speech adaptation effect, with a shift of perceptual boundaries toward the /ada/ category (a tendency to judge ambiguous tokens as /aga/) observed in later blocks. For the /ada/-/ana/ continuum, this speech adaptation effect was asymmetrical, which is interpreted as interacting with a boundary shift toward /ada/ (i.e., ambiguous tokens judged more as /ana/) in the nose-breathing condition. See the online article for the color version of this figure.

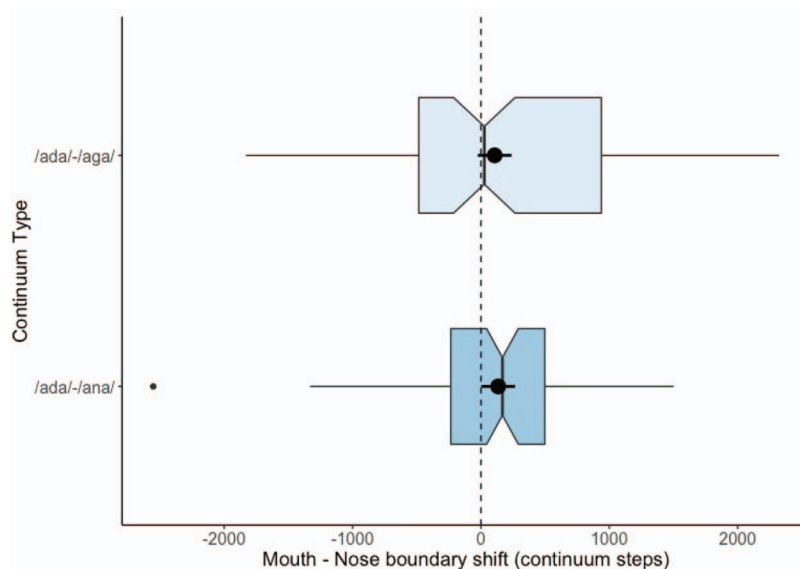


Figure 3. Results here illustrate the boundary shift on each continuum when comparing mouth-breathing minus nose-breathing conditions. Box plots indicate the distribution of the data for all participants ($n = 88$), with notches indicating a distributional estimate of the 95% CIs (see Figure 2). The large dark circles indicate the estimated marginal means from the model (with 95% CIs as error bars on those circles). As can be seen with the overlap between error bars and the dotted line (indicating no shift), there was no significant effect of breathing on the change in estimated perceptual boundaries for the /ada/-/agal/ continuum. In contrast, for the /ada/-/ana/ continuum, there was a significant shift, resulting in numerically larger perceptual boundaries (which means that ambiguous tokens were heard less as /ana/) in the mouth-breathing compared to nose-breathing conditions. See the online article for the color version of this figure.

Altogether, these results show an effect of breathing on the perceptual boundary for the DN continuum, as we had predicted, but not on the DG continuum. This breathing effect also interacted with a strong selective speech adaptation effect, which was visible when examining the other interactions involving Block Order, and which itself interacted with effects of nose- and mouth-breathing in these data.

Discussion

Theories of speech processing have long considered ways that speech perception and production interact. While early motor theories of speech perception had suggested that speech perception was analyzed using gestural information (Galantucci et al., 2006; Liberman & Mattingly, 1985), contemporary accounts of speech perception have taken a more nuanced stance, suggesting that sensorimotor information about gestures is likely helpful in resolving cases of auditory ambiguity, and may be used to establish phonetic categories during acquisition (Hickok, Houde, et al., 2011; Schwartz et al., 2012; Skipper et al., 2017). As we have argued in the introduction, very little is known about the format of the sensorimotor information involved in these interactions.

Our experiments investigated the influence of postural control of the vocal tract articulators on speech perception. We show that keeping one's articulators in a certain position can result in a modulation of ambiguous auditory speech perception. Specifically, we asked whether maintaining a fixed configuration of the velum while breathing through the nose versus the mouth would generate sensorimotor cues that could shift the perceived boundary between

auditory oral and nasal consonants (/ada/ and /ana/). Results showed that sensorimotor cues about postural control (possibly including both efferent cues, and reafferent somatosensory feedback about the location of airflow in the vocal tract) could indeed modulate the perception of ambiguous speech tokens along a continuum.

These findings have broad implications for our understanding of speech processing, particularly for claims about the link between articulatory and auditory targets in speech. Below, we further outline the theoretical implications of this work, then discuss what our results say about the source and structure of sensorimotor influences, and finally propose directions for future work.

Theoretical Implications

Speech processing theories have long featured links between speech perception and production. For example, an early proposal was that articulatory and auditory speech targets were “in parity,” or representationally equivalent (Liberman & Mattingly, 1985). Related theories have suggested that motor elements constitute a part of the mental processing of speech sounds (Browman & Goldstein, 1990, 1992). Other accounts have suggested that auditory targets are distinct from articulatory targets, but that the two interact in important ways (Hickok, Houde, et al., 2011; Rauschecker & Scott, 2009; Schwartz et al., 2012; Skipper et al., 2006), while still others, in contrast, attribute no significant role for sensorimotor information in perception (Diehl et al., 2004). A clearer notion of what kinds of production information are used in perception would help distinguish among these theories.

Previous theories of speech perception have not been very explicit with regard to the formatting of sensorimotor information that is used in perception. One assumption could be that the gestural information used in speech perception is hierarchically organized in ways similar to speech motor control: Broadly speaking, that literature suggests that lower levels of motor control (like the specific movements and trajectories of single articulators) can be separated from higher levels of control (like the dynamic coordination of articulator movements that are required to produce sound; Perkell et al., 2000; Saltzman & Kelso, 1987). Current accounts remain agnostic about how motor information in speech perception is formatted with respect to this hierarchy of motor control: It remains unclear whether it stems primarily from more high-level, coordinative elements, or come from more low-level information about the control of single articulators (like the tongue tip, the lower lip, or the velum). Our results confirm that motor information in speech perception appears not to be encapsulated away from low-level gestural descriptions. Specifically, we show that basic, low-level building blocks of speech gestures—postural control of the velum and the location of airflow while breathing—can modulate the perceptual processing of speech sounds.

At the same time, there may also be limits on the granularity of perceptual predictions from forward models, as we did not find a difference between articulatory configurations where the mouth was open or closed while nose breathing in Experiment 2. Because these distinct mouth positions likely induced subtle differences in both somatosensory feedback (i.e., airflow) as well as the position of the velum (one must maintain a tight seal between the velum and tongue when the mouth is open; Perry, 2011), we might have predicted subtle perceptual differences between these conditions. Our null result here suggests either that there are limits on the resolution of sensorimotor representations in forward models, or that these minor differences were too subtle for our method to detect.

There remains much work to be done in explaining sensorimotor overlap between nonspeech gestures, like respiration, and speech gestures, like articulating /ana/ or /ada/. Future accounts of sensorimotor modulation in the speech domain will have to include a better understanding of what pathways are implicated in the presently observed effects. Nevertheless, our study establishes a new boundary condition with regard to what kind of information is shared between speech processing and nonspeech motor control: Information about the postural control over vocal tract articulators is taken into account when using sensorimotor information in forward models of speech perception.

The Locus of Sensorimotor Influences

Neurocognitive models suggest that sensorimotor modulation of speech perception could arise from two possible sources. The first is derived from the predictions of a forward model, which anticipate the sensory consequences of a planned motor action; the second is derived from the part of the forward model that deals with sensory feedback from reafferent information about an executed vocal gesture (Hickok, 2014; Hickok, Houde, et al., 2011; Schwartz et al., 2012). In the present study, both possible pathways could have contributed to sensorimotor modulation.

The first possibility would claim that the active maintenance of the velum posture (i.e., efferent, or outgoing motor control) may

have triggered a sensory anticipation that was relevant for deciding between either /ada/ or /ana/ in perception. Support for this idea comes from TMS studies (see Möttönen & Watkins, 2012 for a review), where stimulation of a premotor cortex (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007), or even primary motor cortex areas that control a single articulator (D'Ausilio et al., 2009; Möttönen et al., 2013), can result in speech modulation. The current breathing manipulation may indeed have been based on some of the same mechanisms, where velum-related motor activation biased the perception of auditory speech input in ways that are compatible with either nasal or oral sounds.

The second (nonexclusive) possibility is that sensorimotor influences were derived from reafferent (incoming) somatosensory feedback. In the DN condition, for example, reafferent cues about the position of the velum from somatosensation, perhaps in addition to somatosensory cues about the relative amounts of airflow in nasal and oral cavities, may thus have helped to adjudicate an auditory speech task that required the categorization of a sound as either a nasal or oral consonant.

The present results cannot confirm which type of somatosensory information—from velum position or from airflow or both—was more important in modulation, but this could be a target of future work. Sensorimotor influences of this type resemble previous findings, where applying an external source of skin deformation (Ito et al., 2009), had a similar modulatory effect on perception. What makes our findings very different, however, is that the reafferent somatosensory feedback experienced by our experimental participants was not explicitly speech-like. In our study, breathing was not timed to the presentation of the sound, nor did airflow follow the temporal properties of an actual speech gesture, and this lack of temporal consistency had blocked sensorimotor modulation in that previous study. Future research must investigate how that kind of manipulation differs from the type of sensorimotor modulation observed here. For example, one might ask whether externally controlled vocal tract airflow (e.g., by means of an artificial respirator) would similarly be sufficient to trigger modulation of speech perception.

Another important step in identifying the locus of the presently observed effects is to investigate where in the brain these processes occur. Previous studies have described anatomical networks in dorsal regions that may be implicated in integrating sensorimotor and auditory information (Hickok & Poeppel, 2007; Saur et al., 2008; Wilson, Saygin, Sereno, & Iacoboni, 2004), with functional activity reported in many areas including motor and premotor areas, Broca's area, parts of superior temporal cortex, the cerebellum, and areas within parietal lobe (Hickok, 2014; Hickok, Houde, et al., 2011; Skipper et al., 2017). A question raised by our findings is how sensorimotor information about control over a nonspeech posture (i.e., the maintenance of a static position of a single vocal tract articulator, or reafferent somatosensory information about airflow) is transformed within these networks into information about speech sounds.

One intriguing possibility concerns the cerebellum, which has also been suggested as an area of the brain responsible for the synthesis of auditory predictions in forward models (Knolle, Schröger, & Kotz, 2013). This shared functional localization has prompted others to think of the cerebellum as a place where motor planning is integrated with somatosensory information: Here, forward models might generate somatosensory predictions by simu-

lating the execution of a speech gesture, and likewise, reafferent somatosensory feedback to this region might be used to generate predictions about a vocal tract gesture (Hickok, 2012a, 2014). The cerebellum is also implicated in relatively low-level motor control of speech production (Ackermann, Mathiak, & Riecker, 2007), like readjusting the timing of articulator movements to compensate for different speech rates. The cerebellum should certainly be considered as a target for future research exploring the basis of sensorimotor modulation in speech perception.

Future Directions

Our manipulation did not control the force or velocity of airflow, nor the timing of respiration relative to sound presentation, as introducing these controls would have required that participants maintain conscious control over their breathing for a prolonged period. We still find an effect on perception with this experimental variability, but this is perhaps one factor that explains our smaller effect size. Future work is needed to show whether tighter controls on participants' respiration (a methodologically challenging task) would yield a larger effect size, in order to be used in future applications in the clinic. Our results nevertheless represent an important finding for basic researchers in this field, which is that that speech is a specific type of motor gesture, and while "special" in many respects, it may still share mechanisms related to both the perception and production of nonspeech gestures, like breathing, chewing, swallowing, and so forth.

These findings make the prediction that other types of postural control over vocal tract articulators may similarly have a modulatory effect on the perception of speech sounds. For example, holding one's breath at the lips could affect the perception of bilabial plosive consonants, because both induce an elevation of air pressure at the lips. In a similar fashion, hearing vocalizations that share some of the same physical and acoustic features of speech sounds could similarly affect the kinds of speech movements that people make: One possibility is that hearing coughing or throat-clearing with either rounded or unrounded lips may affect the degree of lip-rounding that one produces when speaking.

Another avenue of future research might explore differences between different classes of speech sounds. The articulation of obstruent consonants involves more contact between articulatory landmarks (and thus more somatosensory targets) than vowel articulation does, and as a result is more nonlinear than vowel articulation (Perkell, 2012). Possible differences in the nature of perception–production interactions may arise between consonants and vowels, as somatosensory targets may be less clearly defined for vowels than consonants, and as such, vowels may be less sensitive to modulatory influences from low-level articulator information.

Finally, when considering the format of sensorimotor information in speech processing systems, it is also important to consider evidence from learning and development. Previous studies have suggested, for example, perceptual modulation of infants' earliest speech productions (de Boysson-Bardies & Vihman, 1991) and production-based attentional influences on infants' speech perception (DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014; Streri, Coulon, Marie, & Yeung, 2016). Neuroimaging evidence has also suggested that young infants will activate motor areas of the brain in speech perception, particularly

when listening to non-native speech that is more difficult to process than native speech (Kuhl, Ramírez, Bosseler, Lin, & Imada, 2014). At this young age, the format of sensorimotor information is very unlike adult speech motor control (Steeve, Moore, Green, Reilly, & McMurtrey, 2008), and thus it is unclear how sensorimotor information is structured. The kind of modulation from low-level motor cues tested here may very well be more potent in infancy than adulthood. Intriguingly, evidence for this precise hypothesis comes from studies that manipulate infants' vocal tract configurations. These studies showed modulation of audiovisual speech perception in one case (Yeung & Werker, 2013), and impairment of auditory consonant perception in another (Bruderer, Danielson, Kandhadai, & Werker, 2015; Choi, Bruderer, & Werker, 2019). Such findings suggest an ontogenetic basis of sensorimotor influences on speech perception from low-level articulatory information about nonspeech gestures.

Context of the Research

Both authors' work centers around linkages between speech production and perception and the possible role that forward models may play as a bridge between these aspects of speech. We have worked together on research showing articulatory influences on speech perception. Independently both authors have also worked on how articulatory influences on perception are learned, and involved in other perceptual phenomena, such as perceptual recalibration.

This research grows out of this shared (and joint) work. We have discussed for years how we might address whether subtle differences in the configuration of a perceiver's speech articulators may influence speech perception in line with the simple fact that if forward models are engaged when decoding ambiguous speech (the underlying hypothesis guiding much of our research), the current state of the perceiver's articulators should influence perception as the current position of effectors is necessarily consulted by forward models before issuing a prediction.

We are currently discussing replications and extensions of this research, specifically looking at replicating this effect using new designs, and disentangling the contributions of airflow-related information and somatosensory information about articulator position. This research will involve measurements and/or manipulations of airflow.

References

- Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *The Cerebellum*, 6, 202–213. <http://dx.doi.org/10.1080/14734220701266742>
- Ackermann, H., & Ziegler, W. (2010). Brain mechanisms underlying speech motor control. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *Handbook of phonetics* (2nd ed., pp. 202–250). West Sussex, UK: Wiley-Blackwell. <http://dx.doi.org/10.1002/9781444317251.ch6>
- Amitay, S., Irwin, A., Hawkey, D. J. C., Cowan, J. A., & Moore, D. R. (2006). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners. *The Journal of the Acoustical Society of America*, 119, 1616–1625. <http://dx.doi.org/10.1121/1.2164988>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal*

- of *Memory and Language*, 68, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Bell-Berti, F. (1993). Understanding velic motor control: Studies of segmental context. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, nasalization, and the velum* (pp. 63–85). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-0-12-360380-7.50007-7>
- Berger, C. C., & Ehrsson, H. H. (2013). Mental imagery changes multi-sensory perception. *Current Biology*, 23, 1367–1372. <http://dx.doi.org/10.1016/j.cub.2013.06.012>
- Bishop, D. V., Brown, B. B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *Journal of Speech and Hearing Research*, 33, 210–219. <http://dx.doi.org/10.1044/jshr.3302.210>
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18, 299–320. [http://dx.doi.org/10.1016/S0095-4470\(19\)30376-6](http://dx.doi.org/10.1016/S0095-4470(19)30376-6)
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180. <http://dx.doi.org/10.1159/000261913>
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 13531–13536. <http://dx.doi.org/10.1073/pnas.1508631112>
- Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102, 2360–2370. <http://dx.doi.org/10.1121/1.419620>
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selection in the dorsal auditory stream. *The Journal of Neuroscience*, 33, 5208–5215. <http://dx.doi.org/10.1523/JNEUROSCI.1870-12.2013>
- Choi, D., Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychonomic Bulletin & Review*, 26, 1388–1399. <http://dx.doi.org/10.3758/s13423-019-01601-0>
- Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, 75, 485–491. <http://dx.doi.org/10.2307/1419876>
- Cowie, R., & Douglas-Cowie, E. (1992). *Postlingually acquired deafness: Speech deterioration and the wider consequences* (Vol. 62). New York, NY: Mouton De Gruyter. <http://dx.doi.org/10.1515/9783110869125>
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 381–385. <http://dx.doi.org/10.1016/j.cub.2009.01.017>
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297–319. <http://dx.doi.org/10.1353/lan.1991.0045>
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20120394. <http://dx.doi.org/10.1098/rstb.2012.0394>
- DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior & Development*, 36, 642–649. <http://dx.doi.org/10.1016/j.infbeh.2013.06.007>
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179. <http://dx.doi.org/10.1146/annurev.psych.55.090902.142028>
- Eimas, P. D., Cooper, W. E., & Corbit, J. D. (1973). Some properties of linguistic feature detectors. *Perception & Psychophysics*, 13, 247–252. <http://dx.doi.org/10.3758/BF03214135>
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99–109. [http://dx.doi.org/10.1016/0010-0285\(73\)90006-6](http://dx.doi.org/10.1016/0010-0285(73)90006-6)
- Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, 99, 1161–1173. <http://dx.doi.org/10.1121/1.414884>
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 361–377. <http://dx.doi.org/10.3758/BF03193857>
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504. <http://dx.doi.org/10.1038/nature08572>
- Gick, B., Ikegami, Y., & Derrick, D. (2010). The temporal window of audio-tactile integration in speech perception. *The Journal of the Acoustical Society of America*, 128(5), EL342–EL346. <http://dx.doi.org/10.1121/1.3505759>
- Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, 61, 220–233. <http://dx.doi.org/10.1159/000084159>
- Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., . . . Howard, M. A., III. (2011). Human auditory cortical activation during self-vocalization. *PLoS ONE*, 6, e14744. <http://dx.doi.org/10.1371/journal.pone.0014744>
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25, 408–422. <http://dx.doi.org/10.1016/j.jneuroling.2009.08.006>
- Hickok, G. (2012a). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135–145. <http://dx.doi.org/10.1038/nrn3158>
- 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, 393–402. <http://dx.doi.org/10.1016/j.jcomdis.2012.06.004>
- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29, 2–20. <http://dx.doi.org/10.1080/01690965.2013.834370>
- 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, 214–220. <http://dx.doi.org/10.1016/j.bandl.2011.08.001>
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69, 407–422. <http://dx.doi.org/10.1016/j.neuron.2011.01.019>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. <http://dx.doi.org/10.1038/nrn2113>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216. <http://dx.doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45, 295–310. [http://dx.doi.org/10.1044/1092-4388\(2002/023\)](http://dx.doi.org/10.1044/1092-4388(2002/023))
- House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *The Journal of Speech and Hearing Disorders*, 21, 218–232. <http://dx.doi.org/10.1044/jshd.2102.218>
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1245–1248. <http://dx.doi.org/10.1073/pnas.0810063106>
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26, 235–258. [http://dx.doi.org/10.1016/S0149-7634\(01\)00068-9](http://dx.doi.org/10.1016/S0149-7634(01)00068-9)
- Kawahara, H., Irino, T., & Morise, M. (2011). An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5420–5423). Prague,

- Czech Republic: IEEE. <http://dx.doi.org/10.1109/ICASSP.2011.5947584>
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718–727. [http://dx.doi.org/10.1016/S0959-4388\(99\)00028-8](http://dx.doi.org/10.1016/S0959-4388(99)00028-8)
- Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, 49, 2449–2461. <http://dx.doi.org/10.1016/j.cortex.2012.12.012>
- Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 11238–11245. <http://dx.doi.org/10.1073/pnas.1410963111>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <http://dx.doi.org/10.18637/jss.v082.i13>
- Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, 25, 1325–1336. <http://dx.doi.org/10.1177/0956797614529978>
- Lee, B. S. (1950). Effects of delayed speech feedback. *The Journal of the Acoustical Society of America*, 22, 824–826. <http://dx.doi.org/10.1121/1.1906696>
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63, 1279–1292. <http://dx.doi.org/10.3758/BF03194543>
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-square means (R package version 1.4.4) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13464–13471. <http://dx.doi.org/10.1073/pnas.231459498>
- Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461. <http://dx.doi.org/10.1037/h0020279>
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36. [http://dx.doi.org/10.1016/0010-0277\(85\)90021-6](http://dx.doi.org/10.1016/0010-0277(85)90021-6)
- Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, 10, 179–204. <http://dx.doi.org/10.1080/15475441.2013.829740>
- 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, 1692–1696. <http://dx.doi.org/10.1016/j.cub.2007.08.064>
- Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., & Kondo, T. (2013). Speech misperception: Speaking and seeing interfere differently with hearing. *PLoS ONE*, 8, e68619. <http://dx.doi.org/10.1371/journal.pone.0068619>
- Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, 23, 1190–1197. <http://dx.doi.org/10.1093/cercor/bhs110>
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience*, 29, 9819–9825. <http://dx.doi.org/10.1523/JNEUROSCI.6018-08.2009>
- Möttönen, R., & Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology*, 26, 1103–1118. <http://dx.doi.org/10.1080/02687038.2011.619515>
- Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11, 1217–1222. <http://dx.doi.org/10.1038/nn.2193>
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63, 1–33. <http://dx.doi.org/10.1016/j.cogpsych.2011.05.001>
- Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25, 382–407. <http://dx.doi.org/10.1016/j.jneuroling.2010.02.011>
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., . . . Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28, 233–272. <http://dx.doi.org/10.1006/jpho.2000.0116>
- Perry, J. L. (2011). Anatomy and physiology of the velopharyngeal mechanism. *Seminars in Speech and Language*, 32, 83–92. <http://dx.doi.org/10.1055/s-0031-1277712>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105–110. <http://dx.doi.org/10.1016/j.tics.2006.12.002>
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–360. <http://dx.doi.org/10.1038/nrn2811>
- Ramanarayanan, V., Lammert, A., Goldstein, L., & Narayanan, S. (2014). Are articulatory settings mechanically advantageous for speech motor control? *PLoS ONE*, 9, e104168. <http://dx.doi.org/10.1371/journal.pone.0104168>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724. <http://dx.doi.org/10.1038/nn.2331>
- Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, 18, 6–7. <http://dx.doi.org/10.1111/j.1467-9280.2007.01839.x>
- Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, 94, 84–106. <http://dx.doi.org/10.1037/0033-295X.94.1.84>
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23, 429–435. <http://dx.doi.org/10.1016/j.cogbrainres.2004.11.006>
- Sato, M., Grabski, K., Glenberg, A. M., Brisebois, A., Basirat, A., Ménard, L., & Cattaneo, L. (2011). Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. *Cortex*, 47, 1001–1003. <http://dx.doi.org/10.1016/j.cortex.2011.03.009>
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111, 1–7. <http://dx.doi.org/10.1016/j.bandl.2009.03.002>
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Vry, M., Umarova, R., . . . Kellmeyer, P. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18035–18040. <http://dx.doi.org/10.1073/pnas.0805234105>
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences*, 11, 349–355. <http://dx.doi.org/10.1016/j.tics.2007.06.005>
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25, 336–354. <http://dx.doi.org/10.1016/j.jneuroling.2009.12.004>
- Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, 24, 1824–1830. <http://dx.doi.org/10.1177/0956797613478614>
- Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical*

- Society of America*, 133, EL286–EL292. <http://dx.doi.org/10.1121/1.4794932>
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10, 295–302. <http://dx.doi.org/10.1038/nrn2603>
- Seren, E. (2006). Effect of nasal valve area on inspirator nasal sound spectra. *Otolaryngology—Head and Neck Surgery*, 134, 506–509. <http://dx.doi.org/10.1016/j.otohns.2005.10.038>
- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125, 1103–1113. <http://dx.doi.org/10.1121/1.3058638>
- 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. <http://dx.doi.org/10.1016/j.bandl.2016.10.004>
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101, 260–277. <http://dx.doi.org/10.1016/j.bandl.2007.02.008>
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250–286). Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511541599.009>
- Steeve, R. W., Moore, C. A., Green, J. R., Reilly, K. J., & McMurtrey, J. R. (2008). Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *Journal of Speech, Language, and Hearing Research*, 51, 1390–1404. [http://dx.doi.org/10.1044/1092-4388\(2008/07-0046\)](http://dx.doi.org/10.1044/1092-4388(2008/07-0046))
- Streri, A., Coulon, M., Marie, J., & Yeung, H. H. (2016). Developmental change in infants' detection of visual faces that match auditory vowels. *Infancy*, 21, 177–198. <http://dx.doi.org/10.1111/infa.12104>
- Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE*, 11, e0146813. <http://dx.doi.org/10.1371/journal.pone.0146813>
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26, 952–981. <http://dx.doi.org/10.1080/01690960903498424>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39, 1429–1443. <http://dx.doi.org/10.1016/j.neuroimage.2007.09.054>
- Warren, D. W., Dalston, R. M., & Mayo, R. (1993). Aerodynamics of nasalization. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, nasalization, and the velum* (pp. 119–146). San Diego, CA: Academic Press.
- 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, 701–702. <http://dx.doi.org/10.1038/nn1263>
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880–1882. <http://dx.doi.org/10.1126/science.7569931>
- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, 24, 603–612. <http://dx.doi.org/10.1177/0956797612458802>
- Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 592–597. <http://dx.doi.org/10.1073/pnas.0904774107>

Received March 6, 2018

Revision received July 7, 2020

Accepted August 21, 2020 ■