There's more to speech perception than meets the ear

John F. Houde[1]

Department of Otolaryngology–Head and Neck Surgery, University of California San Francisco, 513 Parnassus Avenue, HSE800, San Francisco, CA 94143

What determines how you perceive the speech sounds you hear? One obvious answer is your ear—your auditory system, because speech is an auditory phenomenon. A less obvious answer is that your mouth—your speech production system—plays a role too. If anything, the direction of influence would seem to go from ear to mouth: children need hearing for normal speech development (1). Yet recent studies, like the one presented in this issue of PNAS by Nasir and Ostry (2), are showing that, surprisingly, your experiences producing speech do indeed affect how you perceive it.

Why Should Speech Production Affect Perception?

In fact, the idea that production experience could affect speech perception has long been used to explain various phenomena in speech perception, like the variations seen in tests of categorical perception. Many speech sounds are determined by peaks in the audio spectrum called formants, and by varying formants, one can create synthetic continua between speech sounds (e.g., /ba/ to /da/, or, as the authors of the article in this issue tested, a continuum from /ɛ/ to /æ/—the vowels in *head* and *had*). When listeners are asked to judge whether the two successive sounds from such continua are the same or different, they are most accurate at a categorical boundary in the sound continuum: this is the boundary observed when the listeners are instead asked to identify the sound they heard from the continuum (e.g., "did you hear /ba/ or /da/?") (3). Categorical boundaries are not only language dependent (e.g., Spanish speakers have no category boundary between /ɛ/ and /æ/ (4)) but also context dependent, with the category boundary between two sounds shifting depending on what sound came before (5).

Context-dependent variations in categorization are necessary because of the variability in produced speech. This is partly due to production constraints—it takes time to move the articulators around, and so speakers do not always reach their intended articulatory targets in running speech (6). It was perhaps natural, therefore, to postulate that listeners use their own past production experience to make allowances for such production constraints as they perceive the speech of others. Such was the intuition behind the Motor Theory of Speech Perception, which postulates that listeners perceive speech from speech sounds by inferring the articulatory goals the speaker had in mind when creating the sounds (7). One version of this theory even posits that an innate cognitive ability evokes articulatory gestures from auditory speech input—that experience with speech is not needed for listeners to have the ability to infer the articulatory goals intended by a speaker (8). A much less controversial theory fits well with a key hypothesized role of the brain: that it works to predict incoming sensory information (9), and does so by using all sources of predictive information available. These sources include the prior history of the stimulus and any past sensory–motor associations experienced. In this sense, the Motor Theory can be restated to say that auditory–motor associations learned from past speech production experience are used, along with all other sources of predictive information, to perceive speech.

Speech production creates associations between motor, somatosensory, and auditory representations.

More recent studies also suggest that production experience affects speech perception. Functional imaging studies have shown that hearing speech results in activity in the speech motor cortex (10), whereas studies using transcranial magnetic stimulation (TMS) find increased excitability in the motor cortex when speech is heard (11). Finally, a very recent study showed that subjects listening to stimuli from an /ɛ/–/æ/ continuum judged ambiguous sounds to be more /æ/-like when their faces were stretched like they were producing /æ/ (12). Yet, in all of these cases, the idea that production experience affects perception has been only a useful explanatory theory: any clear experimental demonstration of production experience affecting perception was lacking.

Finding Evidence that Speech Production Does Affect Perception

Efforts to devise such experiments have been hampered by a key difficulty: in producing speech, you also hear speech (i.e., your own). Thus, any production experiences are also perception experiences, and many experiments have shown that past perceptual experiences affect current perceptions. An example is the selective adaptation effect (13): hearing a speech sound repeated many times affects future perceptions of that sound (e.g., hearing repetitions of /ba/ shifts perception of ambiguous sounds in a /ba/–/da/ continuum to /da/). The first study to avoid such perceptual experience confounds was done by Shiller et al. (14) with the production and perception of voiceless fricatives. The authors used a real-time frequency shifting device to lower the perceived centroid of spectral energy in speakers' productions of /s/ (i.e., the feedback alteration made the /s/ productions sound more /ʃ/-like, as in *she*). Like other speech sensorimotor adaptation studies (15), the authors found this alteration caused speakers to adapt: they raised the centroid of spectral energy of their /s/ productions to make them even more /s/-like. But the authors also conducted speech perception tests of the /s/–/ʃ/ category boundary before and after the adaptation and found the category boundary shifted toward /ʃ/. This boundary shift is opposite the direction expected (i.e., toward /s/) if it was due to perceptual selective adaptation, and, indeed, control subjects producing /s/ with feedback unaltered did show a such a boundary shift toward /s/.

If not caused by selective adaptation, then what mechanism would cause the shift toward /ʃ/? One explanation comes from studies suggesting that during speaking, the brain generates predictions of expected auditory feedback, and that incoming auditory feedback is compared with these predictions (16). Any mismatch between the two indicates an error, but in what system? Most studies have assumed the brain interprets the mismatch as an error in production, but

Author contributions: J.F.H. wrote the paper.

The author declares no conflict of interest.

See companion article on page 20470.

[1]E-mail: houde@phy.ucsf.edu.

is it also possible to interpret the mismatch as an error in the perception system. Thus, experience producing /s/ with audio spectrally shifted down to make /s/ sound like /ʃ/ caused speakers to offset their perception such that the shift seemed smaller. This perceptual offset reduced their perceived need to adapt their production, and it also shifted their /s/–/ʃ/ category boundary toward /ʃ/.

The Shiller et al. study (14) is thus consistent with the idea that production–perception associations guide the perception of future speech sounds. However, the results reported in this issue by Nasir and Ostry (2) are not so easily explained in this way. This study continues a series of studies using a unique system for perturbing subjects' jaw motions (2, 17). The system involves a robotic perturbation device connected to custom-fitted dental appliances that the subjects wear in their mouths. During speaking, the jaw is perturbed in a direction perpendicular to the normal arc of the jaw during speaking—effectively jaw tugs (protrusions) and pushes. The key feature of this perturbation is that it does not alter formant frequencies, something that allowed past studies to show that speech motor control is not concerned solely with acoustic output: despite the perturbations having no effect on formant frequencies, subjects given these jaw perturbations nevertheless adapted their productions to compensate (17). Here, Nasir et al. had subjects adapt to similar tugs: subjects produced words containing /æ/ while their jaws were displaced outwards (protruded) as a function of angular jaw velocity. Consistent with previous findings, after repeated trials, many subjects altered their productions to compensate for the displacement. By the end of training, the angular paths of their jaws were no longer displaced outwardly. Critically, the authors also tested subjects' speech perception before and after the jaw perturbation trials. In the tests, subjects judged the identity of synthesized /hVd/ words (e.g., "head" and "had") where the vowel (V) was drawn from an /ɛ/-/æ/ continuum. Amazingly, the authors found that the jaw perturbation training shifted the subjects' category boundary toward /æ/.

By itself, this perception result would not be surprising because the direction of boundary shift is consistent with a selective adaptation effect acting on auditory perception: repeated productions of /æ/ in the training would also be repeated perceptions of /æ/, which would cause a shift of the /ɛ/–/æ/ boundary toward /æ/. What makes the result remarkable is the fact that the boundary shift was mainly seen in subjects that adapted to the jaw perturbation, not in subjects that did not adapt, and not in subjects run in a control experiment identical to the real experiment but with no jaw perturbations. Because all subjects made the same number of /æ/ productions in the training, all subjects heard the same number of /æ/ repetitions, and thus all subjects would be expected to have similar selective adaptation effects. Why did adapting production make such a difference in whether perception was changed? Unlike the study of Shiller et al. (14), this result is also not easily explained by auditory–motor or auditory–somatosensory associations made during the production training. There would be no obvious benefit to shifting auditory perception: it does nothing to reduce subjects' perceived need to adapt to the somatosensory change. On the other hand, a shift in somatosensory perception that reduced perception of the jaw protrusion would have no obvious effect on auditory perception. Furthermore, if somehow the perceptual adaptation reduced subjects' perceived need to adapt, the authors' regression analysis should have found a negative relationship between motor and perceptual adaptation, with poor motor adaptors being those who chose to adapt their perception more. Instead, the authors found the opposite: the more a subject adapted production, the more he/she adapted perception.

Nevertheless, the results are consistent with the idea that speech production creates associations between motor, somatosensory, and auditory representations, and that altering one of these representations affects the others. The results are also consistent with theories positing a similarity between the cortical systems responsible for speech production and those responsible for performance on speech sound discrimination tests (18). It would be interesting to see, therefore, if the authors' production adaptation also affected the perception of whole words and phrases. However, a specific mechanism to explain the authors' results, as they themselves admit, is lacking. This only makes the result more significant because it challenges our existing models of speech production and perception.

1. Oller DK, Eilers RE (1988) The role of audition in infant babbling. *Child Dev* 59:441–449.
2. Nasir SM, Ostry DJ (2009) Auditory plasticity and speech motor learning. *Proc Natl Acad Sci USA* 106:20470–20475.
3. Liberman AM, et al (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54:358–368.
4. Harris J (1969) *Spanish Phonology* (MIT Press, Cambridge, MA).
5. Repp BH, Liberman AM (1987) Phonetic category boundaries are flexible. *Categorical Perception: The Groundwork of Cognition*, ed Harnad S (Cambridge Univ Press, New York), pp 89–112.
6. Hardcastle WJ, Hewlett N (2006) *Coarticulation: Theory, Data and Techniques* (Cambridge Univ Press, Cambridge, UK).
7. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36.
8. Fowler CA (1996) Listeners do hear sounds, not tongues. *J Acoust Soc Am* 99:1730–1741.
9. Schultz W, Dickinson A (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci* 23:473–500.
10. Wilson SM, et al (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
11. Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
12. Ito T, Tiede M, Ostry DJ (2009) Somatosensory function in speech perception. *Proc Natl Acad Sci USA* 106:1245–1248.
13. Cooper WE (1979) Speech perception and production: Studies in selective adaptation. *Language and Being* (Ablex, Norwood, NJ).
14. Shiller DM, et al (2009) Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am* 125:1103–1113.
15. Houde JF, Jordan IM (1998) Sensorimotor adaptation in speech production. *Science* 279:1213–1216.
16. Houde JF, et al (2002) Modulation of auditory cortex during speech: An MEG study. *J Cogn Neurosci* 14:1125–1138.
17. Tremblay S, Shiller DM, Ostry DJ (2003) Somatosensory basis of speech production. *Nature* 423:866–869.
18. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.