Speech Production: The Force of Your Words

Research on speech production has traditionally focused on how acoustic goals are met. A recent study shows that talking also involves somatosensory goals that do not have any acoustic consequences.

Kevin Munhall

In textbook descriptions of speech production, the different sounds of language are often shown as snapshots of the articulators at a point in time when the sound is being made. This view is illustrated in Figure 1, which shows the vocal tract shapes for three different vowels. It can be seen that complex adjustments of the pharynx, tongue, jaw and lips are required to make these sound distinctions. But this static shape description of speech sounds belies the true intricacy of articulation. Talking has precise temporal demands; a tightly orchestrated sequence of muscle commands is required in order to guide the articulators along elaborate paths within the oral cavity [1]. Dedicated information processing systems, such as a mechanism for managing the auditory feedback of speech [2–4], enable this precise motor guidance. A recent series of studies [5–7] has demonstrated that speech also requires control of the dynamics of movement and processing of the sensory information associated with articulator positions and contacts within the oral cavity. As they recently reported in *Current Biology*, Nasir and Ostry [7] have shown that somatosensory information independently defines one of the planning 'spaces' for the production of speech.

The new work of Nasir and Ostry [7] touches on one of the enduring questions in speech research: with respect to what frame of reference, or coordinate space, is speech planned? Since the earliest X-ray films first revealed the remarkable patterns of motion of the speech articulators (for a review of the X-ray studies in the early 20th century see [8]), there has been a tension between views that give spatial trajectory planning priority [9] versus approaches that maintain that speech goals are framed in acoustic terms [10]. Recently, a more multimodal representation of speech targets has been acknowledged in computational models [11] and empirical studies [12]. In particular, the significance of somatosensory information in the development and maintenance of speech skill is now increasingly recognized [6,13,14].

Nasir and Ostry [7] add to this developing literature by showing that talkers maintain the posture of the jaw during speech in the face of destabilizing force perturbations, even when there are no acoustic consequences to these perturbations. In their study, subjects spoke different words while their mandibles were mechanically linked to a robotic system that delivered perturbations along the lateral axis of the jaw. There is normally little motion along this axis during speech [15] and there were no acoustic consequences of the manipulations. However, their subjects quickly learned to compensate for the perturbations in a manner that maintained the consistency of the lateral jaw position and movement across different speech sounds. The findings, taken together with previous studies from the same lab [5], show that there are aspects of speech movements that talkers want to get right even when there is no apparent need to do so from an acoustical point of view.

Afferent information from the somatosensory system is the basis for these adjustments, and similar sensorimotor processing is presumably responsible for a range of talking 'feats'. We are capable of talking with a pen or pipe in our mouth and, though this is less socially acceptable, also with a mouthful of food or while nervously chewing on a fingernail. These talking conditions clearly necessitate detailed somatosensory involvement. Nasir and Ostry [7] argue that this type of sensory processing is always involved in speech production, and that it is at the heart of the motor precision we attribute to speech.

To maintain precision when dealing with the dynamical interactions that are encountered in natural movements, talkers use different compensation strategies. In the study by Nasir and Ostry [7], subjects modified the impedance of the jaw; the increased stiffness of the joint was sufficient to counteract lateral forces generated by the robot. However, other novel force conditions are known

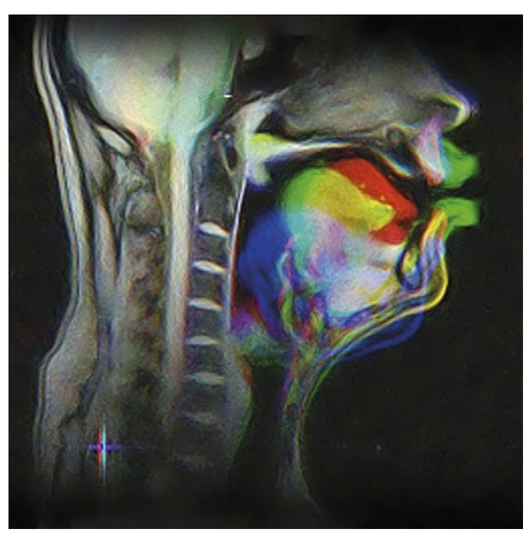

Figure 1. Superimposed mid-saggital images of the vocal tract during the production of the vowels /i/ (red), /u/ (green) and /a/ (blue). (Illustration created by Mark Tiede.)

to induce the generation of precise compensatory forces that counteract the unexpected loads [6]. What determines the relative contributions of these two types of response to disturbances is not completely understood, but work on limb control suggests that the balance of responses might vary over the course of learning. An initial response is to stiffen in the face of unfamiliar forces but then, if the disturbances are predictable, the CNS quickly also learns to produce opposing forces to counteract the disturbing force [16]. This learning can happen very rapidly with crude predictive responses seen even on the second trial [17].

The picture that is emerging of speech is one of an activity that is produced with the aid of rich, predictive motor representations that learn aspects of the task, performance environment and the feedback that is expected when an action is performed. For speech, this means that the sound system of the language, the individual vocal tract morphology, and the expected auditory and somatosensory feedback are contained in a motor representation that is drawn upon when a sound sequence is planned. In its reliance on these learned representations, speech motor control shows marked similarities to limb control [18].

In spite of the linguistic message being conveyed and the higher order influences on the form of this message, speech is in the end a motor activity; as Nasir and Ostry [7] report, this requires that the mechanical environment and dynamic stability of the final filter in the language production process, the speech musculature, be taken into account.

References

1. Perkell, J.S. (1969). Physiology of Speech Production (Cambridge, MA: M.I.T. Press).
2. Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science *279*, 1213–1216.
3. Jones, J.A., and Munhall, K.G. (2005). Remapping auditory-motor representations in voice production. Curr. Biol. *15*, 1768–1772.
4. Purcell, D.W., and Munhall, K.G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. J. Acoust. Soc. Am. *120*, 966–977.
5. Tremblay, S., Shiller, D.M., and Ostry, D.J. (2003). Somatosensory basis of speech production. Nature *423*, 866–869.
6. Ito, T., Gomi, H., and Honda, M. (2004). Dynamical simulation of speech cooperative articulation by muscle linkages. Biol. Cybernet. *91*, 275–282.
7. Nasir, S., and Ostry, D.J. (2006). Somatosensory precision in speech production. Curr. Biol. *16*, 1918–1923.
8. Moll, K. (1960). Cinefluorgraphic techniques in speech research. J. Speech Hear. Res. *3*, 227–241.
9. Goldstein, L., and Fowler, C.A. (2003). Articulatory phonology: a phonology for public language use. In Phonetics and Phonology in Language Comprehension and Production, N. Schiller and A. Meyer, eds. (Berlin: Mouton de Gruyter), pp. 159–208.
10. Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. J. Phonet. *28*, 233–272.
11. Guenther, F.H., Hampson, M., and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychol. Rev. *105*, 611–633.
12. Perkell, J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F.H. (2004). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. J. Speech Lang. Hear. Res. *47*, 259–269.
13. Shiller, D.M., Laboissiere, R., and Ostry, D.G. (2002). Relationship between jaw stiffness and kinematic variability in speech. J. Neurophysiol. *88*, 2329–2340.
14. Shiller, D.M., Houle, G., and Ostry, D.G. (2005). Voluntary control of human jaw stiffness. J. Neurophysiol. *94*, 2207–2217.
15. Ostry, D.J., Vatikiotis-Bateson, E., and Gribble, P.L. (1997). An examination of the degrees of freedom of human jaw motion in speech and mastication. J. Speech Lang. Hear. Res. *40*, 1341–1351.
16. Shadmehr, R., and Mussa-Ivaldi, F.A. (1994). Adaptive representation of dynamics during learning of a motor task. J. Neurosci. *14*, 3208–3224.
17. Milner, T., and Franklin, D. (2005). Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. J. Physiol. *567*, 651–664.
18. Tin, C., and Poon, C.S. (2005). Internal models in sensorimotor integration: Perspectives from adaptive control theory. J. Neural Eng. *2*, S147–S163.

Department of Psychology, Department of Otolaryngology, Queen's University, Kingston, Ontario, Canada, K7L 3N6.
E-mail: Kevin.Munhall@queensu.ca

DOI: 10.1016/j.cub.2006.09.049

Meiotic Spindle: Sculpted by Severing

Katanin is a conserved AAA ATPase with the ability to sever microtubules, but its biological function in animal cells has been obscure. A recent study using electron tomography has found that katanin stimulates the production of microtubules in the meiotic spindles of *Caenorhabditis elegans* oocytes.

Katharina Ribbeck and Timothy J. Mitchison

The protein katanin earned its name by its ability to sever microtubules — katanin is the Japanese word for sword. It was discovered by biochemical purification from sea urchin embryos [1] and is widely conserved in animals and plants, but its biological functions have been difficult to elucidate. It is not obvious why cells need a microtubule-severing enzyme; microtubules mediate transport by providing long tracks for motor proteins to travel along, and they can grow and shrink rapidly from their ends, which should allow remodeling of the cytoskeleton without the need for severing. Experiments in neurons and plant cells have suggested that katanin trims microtubules into fragments that can be efficiently transported and orientated within the cytoplasm, facilitating morphogenesis of complex cytoskeletons [2–4]. These experiments used interphase cells, and did not explain an early observation that katanin activity is highly increased in *Xenopus* egg extract in M-phase relative to interphase, which pointed toward a mitotic function of katanin, such as rapid disassembly of interphase