# Modularizing speech

*Bryan Gick[1,2]\* and Ian Stavness[3]*

[1] Department of Linguistics, University of British Columbia, Vancouver, BC, Canada
[2] Haskins Laboratories, New Haven, CT, USA
[3] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
*Correspondence: gick@mail.ubc.ca

The need to reduce the dimensionality of movement systems, and thereby to decrease cognitive load, has long been recognized as a central challenge for theories of motor control (Bernstein, 1967). A large body of work in neurophysiology, biomechanics, and computation has substantiated the view that control of body movements is distributed among a manageable number of degrees of freedom corresponding to neuromuscular *modules* (e.g., Bizzi et al., 1991), or proportionally fixed groupings of muscles (see e.g., Ting et al., 2012 for a recent review). Current work in computational neuroscience provides evidence that the nervous system uses such modules to achieve dimensionality reduction (e.g., Berger et al., 2013). It is our opinion that a fully realized modular approach to speech movement will have a profound impact on models of speech.

In speech-related fields, researchers had begun formulating ideas for modularizing speech movements even prior to Bernstein's influence. Cooper et al. (1958), for instance, in proposing their notion of the "action plan," described for speech an inventory of muscle activations not unlike Bernstein's "muscle synergies": "we may hope to describe speech events in terms of a rather limited number of muscle groups…" (p. 939). Later, Turvey (1977) adopted the term *coordinative structure* to refer to similar neuromuscular groupings. Easton (1972) had first defined coordinative structures as neuromuscular organizations "underlying all volitionally composed movements… activated by a single command," such that "the CNS [central nervous system] may be said to have at its disposal a library, or set, of these responses" (p. 591).

However, Turvey et al. (1978) shifted focus away from neurophysiology, observing that coordinative structures are "formally equivalent" to tasks in control space (1978, p. 566). Subsequent speech researchers have taken this lead, focusing on developing models of control space (e.g., Kelso et al., 1986a; Tourville and Guenther, 2011), with little or no attention given to modeling the neurophysiology of embodied speech.

Meanwhile, researchers in other areas have built a substantial volume of experimental and modeling research around the neuromuscular organization and biomechanics of non-speech movement, including work on complex fine motor systems such as the fingers (e.g., Overduin et al., 2012) and eyes (e.g., Wei et al., 2010). However, speech, along with many other functions of the upper vocal tract, has remained a conspicuous omission from the literature on neuromuscular modularization. This omission may be ascribed at least in part to the relatively greater complexity of both the muscular structures (e.g., Sanders and Mu, 2013) and the multidimensional control space (e.g., Houde and Jordan, 1998; Tremblay et al., 2003; Gick and Derrick, 2009; Ghosh et al., 2010; Perkell, 2012) of speech. Kelso et al. (1986b) describe this position clearly, stating that mapping their control paradigm onto "real" body structures is "not feasible for the speech articulators whose peripheral biomechanics are much more complex (than upper limbs), e.g., the passive tissue properties and muscular forces of the tongue and lips."

The great majority of evidence for modularization derives from experiments on non-human spinal structures (see Tresch et al., 2002) and from direct recordings of neuromuscular activity using electromyography (see Kutch and Valero-Cuevas, 2012). However, neither of these methods is likely to be as effective for understanding neural control of speech, first because upper airway innervation is predominantly cranial rather than spinal, and second because of the known challenges of experimentally recording comprehensive or even representative neuromuscular activity from EMG, even in less complex tasks than speech (Pittman and Bailey, 2009) and in comparatively less complex neuromuscular systems (Hug, 2011; De Rugy et al., 2013). Because of this, we anticipate that biomechanics will necessarily play a more central role in accessing the modular neuromuscular structures that underlie speech production.

In our view, neuromuscular modules are built specifically to drive body structures that are biomechanically efficacious, enabling them to operate feed-forward, i.e., with little or no central feedback control. This has often been assumed as a premise underlying modularization (e.g., Loeb et al., 2000; d'Avella et al., 2003; Loeb, 2012), but has seldom been tested (see Berniker et al., 2009 for a rare exception), and never applied to speech. Recent advances in modeling speech biomechanics (e.g., Nazari et al., 2011; Stavness et al., 2012a,b) have enabled our group to begin identifying some of the biomechanical properties that we consider to be the hallmarks of speech production modules, most notably pervasive saturation effects that enable feed-forward control of speech structures (Gick et al., in press). At least some of these biomechanically optimized speech production modules correspond well with speech "gestures," long

described as movement-related primitives of speech (e.g., Browman and Goldstein, 1986).

While there remains some controversy around whether these modules are best defined in terms of their neural (e.g., d'Avella and Bizzi, 2005; Safavynia and Ting, 2013), biomechanical (Dominici et al., 2011; Kutch and Valero-Cuevas, 2012), or computational (Todorov, 2004; Diedrichsen et al., 2010; Loeb, 2012; De Rugy et al., 2013) properties, all of these aspects of control will be necessary components of a complete theory (see Bizzi and Cheung, 2013), and at present none of these aspects have been well explored for speech and upper airway control.

Developing a theory of speech production that accords with current work on neuromuscular modularization, we believe, has the potential to link a number of fields and methodologies surrounding a central question in cognitive science, with implications for all aspects of speech research, from phonetics and phonology to the phylogenetic and ontogenetic development of speech. In addition to bringing another complex motor system into the broader discussion of neural modules, modularizing speech at the neuromuscular level promises a major advance for speech models, constituting a "missing link" between speech movement primitives (Ramanarayanan et al., 2013) and newly discovered cortical regions associated with speech production (Bouchard et al., 2013).

## ACKNOWLEDGEMENT

## REFERENCES

Berger, D. J., Gentner, R., Edmunds, T., Pai, D. K., and d'Avella, A. (2013) Differences in adaptation rates after virtual surgeries provide direct evidence for modularity. *J. Neurosci.* 33, 12384–12394. doi: 10.1523/JNEUROSCI.0122-13.2013

Berniker, M., Jarc, A., Bizzi, E., and Tresch, M. C. (2009). Simplified and effective motor control based on muscle synergies to exploit musculoskeletal dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7601–7606. doi: 10.1073/pnas.0901512106

Bernstein, N. (1967). *The Coordination and Regulation of Movements.* 1st English Edn, New York, NY: Pergamon Pr.

Bizzi, E. and Cheung, V. C. K. (2013). The neural origin of muscle synergies. *Front. Comput. Neurosci.* 7:51. doi: 10.3389/fncom.2013.00051

Bizzi, E., Mussa-Ivaldi, F. A., and Giszter, S. (1991). Computations underlying the execution of movement: a biological perspective. *Science* 253, 287–291. doi: 10.1126/science.1857964

Bouchard, K. E., Mesgarani, N., Johnson, K. and Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327–332. doi: 10.1038/nature11911

Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonol. Yearb.* 3, 219–252. doi: 10.1017/S0952675700000658

Cooper, F. S., Liberman, A. M., Harris, K. S., and Grubb, P. M. (1958). "Some input-output relations observed in experiments on the perception of speech," in *Proceedings of the 2nd International Congress on Cybernetics*, (Namur), 930–941.

d'Avella, A., and Bizzi, E. (2005). Shared and specific muscle synergies in natural motor behaviors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3076–3081. doi: 10.1073/pnas.0500199102

d'Avella, A., Saltiel, P., and Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.* 6, 300–308. doi: 10.1038/nn1010

De Rugy, A., Loeb, G. E., and Carroll, T. J. (2013). Are muscle synergies useful for neural control? *Front. Comput. Neurosci.* 7:19. doi: 10.3389/fncom.2013.00019

Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2010). The coordination of movement: optimal feedback control and beyond. *Trends Cogn. Sci.* 14, 31–39. doi: 10.1016/j.tics.2009.11.004

Dominici, N., Ivanenko, Y. P., Cappellini, G., d'Avella, A., Mondi, V., Cicchese, M., et al. (2011). Locomotor primitives in newborn babies and their development. *Science* 334, 997–999. doi: 10.1126/science.1210617

Easton, T. A. (1972). On the normal use of reflexes. *Am. Sci.* 60, 591–599.

Ghosh, S., Matthies, M., Maas, E., Hanson, A., Tiede, M., Ménard, L., et al. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *J. Acoust. Soc. Am.* 128, 3079–3087. doi: 10.1121/1.3493430

Gick, B., Anderson, P., Chen, H., Chiu, C., Kwon, H. B., Stavness, I., et al. (in press). Speech function of the oropharyngeal isthmus: a modeling study. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572

Houde, J. F. and Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science* 279, 1213–1216. doi: 10.1126/science.279.5354.1213

Hug, F. (2011). Can muscle coordination be precisely studied by surface electromyography? *J. Electromyogr. Kinesiol.* 21, 1–12. doi: 10.1016/j.jelekin.2010.08.009

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (1986a). The dynamical perspective on speech production: data and theory. *J. Phon.* 14, 29–59.

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (1986b). Intentional contents, communicative context, and task dynamics: a reply to the commentators. *J. Phon.* 14, 171–196.

Kutch, J. J. and Valero-Cuevas, F. J. (2012). Challenges and new approaches to proving the existence of muscle synergies of neural origin. *PLoS Comput. Biol.* 8:e1002434. doi: 10.1371/journal.pcbi.1002434

Loeb, D. E. (2012). Optimal isn't good enough. *Biol. Cybernet.* 106, 757–765. doi: 10.1007/s00422-012-0514-6

Loeb, D. E., Giszter, S. F., Saltiel, P., Mussa-Ivaldi, F. A. and Bizzi, E. (2000). Output units of motor behavior: an experimental and modeling study. *J. Cogn. Neurosci.* 12, 78–97. doi: 10.1162/08989290051137611

Nazari, M. A., Perrier, P., Chabanas, M., and Payan, Y. (2011). Shaping by stiffening: a modeling study for lips. *Mot. Control* 15, 141–168.

Overduin, S. A., d'Avella, A., Carmena, J. M., and Bizzi, E. (2012). Microstimulation activates a handful of muscle synergies. *Neuron* 76, 1071–1077. doi: 10.1016/j.neuron.2012.10.018

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neurolinguist.* 25, 382–407. doi: 10.1016/j.jneuroling.2010.02.011

Pittman, L. J., and Bailey, E. F. (2009). Genioglossus and intrinsic electromyographic activities in impeded and unimpeded protrusion tasks. *J. Neurophysiol.* 101, 276–282. doi: 10.1152/jn.91065.2008

Ramanarayanan, V., Goldstein, L., and Narayanan, S. S. (2013). Articulatory movement primitives – extraction, interpretation and validation. *J. Acoust. Soc. Am.* 134, 1378–1394. doi: 10.1121/1.4812765

Safavynia, S. A. and Ting, L. H. (2013). Sensorimotor feedback based on task-relevant error robustly predicts temporal recruitment and multidirectional tuning of muscle synergies. *J. Neurophysiol.* 109, 31–45. doi: 10.1152/jn.00684.2012

Sanders, I. and Mu, L. (2013). A three-dimensional atlas of human tongue muscles. *Anat. Rec.* 296, 1102–1114. doi: 10.1002/ar.22711

Stavness, I., Lloyd, J. E., and Fels, S. S. (2012a). Automatic prediction of tongue muscle activations using a finite element model. *J. Biomech.* 45, 2841–2848. doi: 10.1016/j.jbiomech.2012.08.031

Stavness, I., Gick, B., Derrick, D., and Fels, S. S. (2012b). Biomechanical modeling of english /ɹ/ variants. *J. Acoust. Soc. Am. Express Lett.* 131, 355–360. doi: 10.1121/1.3695407

Ting, L. H., Chvatal, S. A., Safavynia, S. A., and McKay, J. L. (2012). Review and perspective: neuromechanical considerations for predicting muscle activation patterns for movement. *Int. J. Numer. Methods Biomed. Eng.* 28, 1003–1014. doi: 10.1002/cnm.2485

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci.* 7, 907–915. doi: 10.1038/nn1309

Tourville, J. A., and Guenther, F. H. (2011). The DIVA model: a neural theory of speech acquisition and production. *Lang. Cogn. Processes* 26, 952–981. doi: 10.1080/01690960903498424

Tremblay, S., Shiller, D. M., and Ostry, D. (2003). Somatosensory basis of speech production. *Nature* 423, 866–869. doi: 10.1038/nature01710

Tresch, M. C., Saltiel, P., d'Avella, A., and Bizzi, E. (2002). Coordination and localization

in spinal motor systems. *Brain Res. Rev.* 40, 66–79. doi: 10.1016/S0165-0173(02)00189-3

Turvey, M. T. (1977). "Preliminaries to a theory of action with reference to vision," in *Perceiving, Acting and Knowing: Toward all Ecological Psychology,* eds R. Shaw and J. Bransford (Hillsdale, NJ: Lawrence Erlbaum Associates), 211–265.

Turvey, M. T., Shaw, R. E., and Mace, W. M. (1978). "Issues in a theory of action: degrees of freedom, coordinative structures, and coalitions, in *Attention and Performance, Vll.* ed J. Requin (Hillsdale, NJ: Lawrence Erlbaum), 557–595.

Wei, Q., Sueda, S., and Pai, D. K. (2010). Physically-based modeling and simulation of extraocular muscles. *Prog. Biophys. Mol. Biol.* 103, 273–283. doi: 10.1016/j.pbiomolbio.2010.09.002