The quantal larynx in action:
Smooth and abrupt aspects of laryngeal motion observed in laryngoscopic videos

Scott Reid Moisik, Bryan Gick, & John H. Esling

Stevens (1989) theorized that phonetic (and, by implication, phonological) categories of speech sound systems are grounded in and shaped by certain non-linear relations found in the acoustic-articulatory and auditory-acoustic domains. Speech biomechanics was suggested at the time of Stevens' publication to act in a similar way (Fujimura, 1989), and recent computational modeling work demonstrates that many different parts of the vocal tract exhibit articulatory-configurational stabilities under the activation of specific combinations of muscles (e.g., Nazari, Perrier, Chabanas, & Payan, 2011; Gick, Stavness, Chiu, & Fels, 2011; Gick et al., 2014).

Recently we added the larynx to this list of structures that can be considered to be capable of such "quantal" articulatory-biomechanical behaviour (Moisik & Gick, 2017). This work, which draws on the phonetic postural space elaborated by Esling and colleagues (Esling & Harris, 2003, 2005; Esling, 2005; Edmondson & Esling, 2006; inter alia), identifies intercorniculate (consistent with Fink, 1974), vocal-ventricular fold, and cuneiform-epiglottal contacts as key stricture events that characterize the degree of stability achievable for any given posture and constitute "saturation points" within the functional articulatory-phonetic space of the larynx.

Together, such work marks a shift away from the agonist-antagonist view of the muscular action in motor control, in which laryngeal actions are seen as smoothly varying in relation to adjustments in muscle contraction, towards the possibility that speech articulation should unfold in rather abrupt steps (rather than smoothly varying transitions) from one speech sound to the next. This prediction seems consistent with a view that phonological categories, being part of the discrete code of speech, would benefit from actual physical implementation as semi-discrete events, at least in part, the most relevant being the segmental (consonant and vowel) layer of the speech signal.

In this talk, we follow up on the predictions arising from our laryngeal biomechanical modeling work by examining a collection of laryngoscopic videos produced by John Esling as part of research which originally intended to expand on the taxonomy of phonetic laryngeal gestures and which made significant discoveries about the production of pharyngeal and epiglottal sounds (Esling, 1996). These videos feature cardinal productions of these sounds contextualized within intervocalic [i] context and the English carrier sentence ("He hit i_i quickly"). Because the videos feature excellent coverage of those sounds examined in the biomechanical modeling work (Moisik & Gick, 2017), a firm basis exists for comparison between the real productions and the quantality evaluated by means of the model.

The videos were analyzed using image analysis techniques, including optical flow (Horn & Schunck, 1981; Moisik, Lin, & Esling, 2014) and non-negative matrix factorization (Lee & Seung, 1999; Ramanarayanan, Goldstein, & Narayanan, 2013), in relation to the segmental boundary information and pitch (f0). Transitions within these postural signals were evaluated by means of the Q-score measure proposed in (Moisik & Gick, 2017) as a means to evaluate objectively the degree of stability of a given "quantal effect". Our preliminary results indicate that while there are clear indications of rapid stabilization motions of the larynx during the production of a range of glottal and aryepiglotto-epiglottal speech sounds ([h], [ʔ], [ʡ], [ʕ], [ħ], [ʢ], and [ʜ]) and sounds dependent on laryngeal action (aspiration and the modal phonatory posture), we also observe that (vertical laryngeal) movements correlated with pitch show a more smoothly varying nature, consistent with the relatively long-term nature and non-discrete (and expressive) functional demands of the intonation contour (which was mostly of a declarative type in these videos). Also somewhat surprising was the enhanced degree of laryngeal closure (typically at the level of aryepiglotto-epiglottal stricture) observed in relation to prosodic-morphological boundary marking. It is not clear how idiosyncratic such closure is, but it is consistent with early observations by Lindqvist-Gauffin (1972). It leads us to the suspicion that such extreme closure may not be all that extreme and in fact perhaps rather more common in ordinary speech than one might suppose.

This work helps validate the discoveries of our biomechanical modeling work and provides further evidence that speech production is embedded within a system that has a natural capacity for supporting the discrete segmental units of the phonetic and phonological code. The significance of this is that it supports the view first implicated by Stevens's work that phonology, thought to be a system of discrete and abstract cognitive units, is embodied by physical systems which ordinarily exhibit semi-discrete tendencies. This further leads to the possibility that phonology is scaffolded upon and thereby benefits from and is reinforced by such tendencies.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, *23*(2), 157–191.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *JIPA*, *26*(2), 65–88.

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics*, *50*, 13–44.

Esling, J. H., & Harris, J. G. (2003). An expanded taxonomy of states of the glottis. In *ICPhS 15* (Vol. 1, pp. 1049–1052). Barcelona, Spain.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 347–383). Mahwah, New Jersey: Erlbaum.

Fink, B. R. (1974). Spring mechanisms in the human larynx. *Acta Oto-Lar.*, *77*(1–6), 295–304.

Fujimura, O. (1989). Comments on "On the Quantal Nature of Speech", by K. N. Stevens. *Journal of Phonetics*, *17*, 87–90.

Gick, B., Anderson, P., Chen, H., Chiu, C., Kwon, H. B., Stavness, I., … Fels, S. (2014). Speech function of the oropharyngeal isthmus: a modelling study. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *2*(4), 217–222.

Gick, B., Stavness, I., Chiu, C., & Fels, S. S. (2011). Categorical variation in lip posture is determined by quantal biomechanical-articulatory relations. *Canadian Acoustics*, *39*(3), 178–179.

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*(1), 185–203.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565

Lindqvist-Gauffin, J. (1972). A descriptive model of laryngeal articulation in speech. *QPSR-KTH 13*(2–3), 1–9.

Moisik, S. R., & Gick, B. (2017). The Quantal Larynx: The Stable Regions of Laryngeal Biomechanics and Implications for Speech Production. *Journal of Speech Language and Hearing Research*, *60*(3), 540–560. https://doi.org/10.1044/2016_JSLHR-S-16-0019

Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *JIPA*, *44*(01), 21–58.

Nazari, M. A., Perrier, P., Chabanas, M., & Payan, Y. (2011). Shaping by stiffening: a modeling study for lips. *Motor Control*, *15*(1), 141–168.

Ramanarayanan, V., Goldstein, L., & Narayanan, S. S. (2013). Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation. *The Journal of the Acoustical Society of America*, *134*(2), 1378–1394. https://doi.org/10.1121/1.4812765

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.