Analysis of laryngoscopic videos using Nonnegative Matrix Factorization (NMF): Characterizing the quantality of laryngeal articulations

Scott R. Moisik[1], John H. Esling[2], & Bryan Gick[3]
[1]Nanyang Technological University, Singapore, [2]The University of Victoria, Canada,
[3]University of British Columbia, Canada
scott.moisik@ntu.edu.sg, esling@uvic.ca, gick@mail.ubc.ca

Following the work of Stevens' (1989) quantal theory but in the context of speech biomechanics (e.g., Gick, Stavness, Chiu, & Fels, 2011; Gick et al., 2014), Moisik and Gick (2017) presented a three-dimensional computer model of the larynx, called QL2, (developed in ArtiSynth; www.artisynth.org) which characterizes the quantal nature of laryngeal articulations, ranging in constriction from glottal fricative (expiration) [h] to epiglottal stop [ʡ]. This work found that the more tissue contact involved (e.g., intercorniculate, vocal-ventricular fold, and cuneiform-epiglottal), the more "quantal" (read stable) the articulation and the more likely one would observe an articulation of that sort. For instance, glottal stop without ventricular fold reinforcement is not as stable, according to QL2, as glottal stop with reinforcement, so we would expect reinforced glottal stop to occur more often. The model thus makes predictions about the type of motion we might expect of these different articulations, and our purpose here is to see if these predictions bear out in actual laryngoscopic data.

To see if these results apply to actual laryngeal articulation, we use a novel approach that makes use of nonnegative matrix factorization (NMF; see Lee & Seung, 1999) and a means to quantify quantality called the Q-score. First, we present our validation of the Q-score measure to demonstrate its utility in distinguishing smooth articulations from abrupt ones. We then show how NMF analysis extracts articulatory states as factors (or "parts") of laryngoscopic videos of [h], [ʔ], [ħ], [ʕ], and [ʡ] (in a carrier sentence context). Analysis of the Q-score values of the mixture/encoding signals for these state factors in relation to segmental analysis of the video presents us with a means to characterize the quantality of individual segments.

Our results are somewhat at odds with the predictions of the model, but suggest that [h] in fact is the most quantal (but this may be general to the consonants with vocal fold abduction). We also observe, however, that glottal stop, which is not very quantal by our measure, is always produced with ventricular reinforcement yet actually also tends to be produced with progressive aryepiglotto-epiglottal narrowing. This work has implications for helping to understand those quasi-discrete aspects of speech production that may underlie the categorical nature of phonological structure.

References

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.

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

Moisik, S. R., & Gick, B. (2017). The quantal larynx: the stable regions of laryngeal biomechanics and implications for speech production. *JSLHR*, *60*(3), 540–560.

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