AGENT MODEL REVEALS THE INFLUENCE OF VOCAL TRACT ANATOMY ON SPEECH DURING ONTOGENY AND GLOSSOGENY

Rick Janssen[1], Scott R. Moisik[2,1], and Dan Dediu[*3,1]

[*]Corresponding Author: Dan.Dediu@mpi.nl
[1]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Linguistics and Multilingual Studies, Nanyang Technological University, Singapore
[3]Collegium de Lyon, Institut d'Études Avancées, Lyon, France

This study introduces a new model for the investigation of the complex manner in which vocal tract anatomy affects human speech production and may influence language change and evolution. The anatomy of the human vocal tract has long been recognized to play a crucial role in speech production and patterning (Fant, 1971; Ohala, 1983). It imposes discrete relations between articulatory parameters and acoustics (Stevens & Keyser, 2010), with highly nonlinear mappings between them (Stevens, 1968, 1989), and it has been recently suggested that inter-individual and patterned inter-population variation in the anatomy of the vocal tract might play a role in explaining patterns of linguistic diversity (Dediu, Janssen, & Moisik, 2017).

We investigate these complex relationships by instructing a computer-simulated agent to learn to reproduce, as well as possible, target speech sounds by controlling the articulators of a detailed 3D geometric model of the human vocal tract based on the VocalTractLab 2.1 (Birkholz, Jackèl, & Kroger, 2006), modified to allow changes in larynx height and hard palate shape. More precisely, the agent minimizes the Euclidean distance (in the F1–F5 formant space) between the target and the produced sounds using a genetic algorithm that optimizes the synaptic weights of a neural network that maps formants to articulatory parameter values[1]. Here, we apply this model to two case studies, both using the five-vowel system [a], [æ], [i], [u], and [ə], but investigating the effects of variation in different components of the vocal tract.

In the first case study, we revisit the debate concerning the role of *larynx height* in human speech, which has important implications for the evolution of speech

---

[1]Unlike other studies (Guenther, 2006; Kröger, Kannampuzha, & Neuschaefer-Rube, 2009), we focus on the anatomy and not on neuro-developmental effects, resulting in a cognitive architecture deliberately designed to be domain-general and based on well-established machine-learning algorithms; however, our architecture is modular and other learning mechanisms can be plugged-in.

and language (Fitch, Boer, Mathur, & Ghazanfar, 2016; Boë et al., 2013; Lieberman, 2012). Our model generally agrees with the conclusions of Boë, Heim, Honda, and Maeda (2002) that a descended larynx is not a necessary prerequisite for modern human speech, but also highlights that there seems to be an optimal larynx height for vowel production approximating that of a modern human female (supporting De Boer, 2010), with a lower or higher larynx seemingly less well suited to produce the optimally expressive vowel inventories of modern human languages. Our model also allows the investigation of the behavior of the other articulators, and we found that certain articulators (such as the tongue and lips) play an important role in the (imperfect) compensation of larynx height, and that they do not seem to act individually but as components of rather complex articulatory subsystems.

The second case study looked into the more subtle effects of the shape of the *hard palate*. The hard palate affects the articulatory gestures required to produce a large set of speech sounds, including [ɹ] (Tiede, Boyce, Holland, & Choe, 2004; Zhou, Espy-Wilson, Tiede, & Boyce, 2007; Tiede, Boyce, Espy-Wilson, & Gracco, 20010), sibilants (Weirich & Fuchs, 2011), and high vowels (Mooshammer, Perrier, Geng, & Pape, 2004; Brunner, Fuchs, & Perrier, 2005, 2009). As in the previous case, we first investigated the systematic differences between the acquired and target vowel system in individual agents. However, while larynx height has a relatively large effect, hard palate shape shows very weak and subtle effects. This prompted us to use *iterated learning* in chains of agents to investigate the amplification of anatomical biases (Kirby & Hurford, 2002; Kirby, Dowman, & Griffiths, 2007). More precisely, we tested five chains (in a given chain all agents have the same hard palate anatomy; the five hard palate antomies are: an artifically "low" and an artificially "high" configuration, the "standard" configuration of our model, as well as the palate shapes of two human participants acquired with MRI); each chain was run for 50 generations, starting with the same five target vowels, and we ran 20 independent replications. We found that iterated transmission significantly affects the acoustics and articulation across generations, with most vowels and chains reaching a plateau before 40 generations. We found differences between the five palate shapes (including those of real participants), with a very slight tendency for the vowels to become more similar to each other (see Smith, Tamariz, & Kirby, 2013) on a background of vowel- and hard palate shape-specific effects.

In conclusion, using a detailed geometric model of the vocal tract, coupled with modern machine learning algorithms, allows the precise investigation of subtle effects of anatomical variation on speech (anatomical biases). The study showed human larynx height exerts strong biases that are expressed *ontogenetically* (i.e., during an individual's language acquisition). Weaker biases from the hard palate shape are mainly expressed *glossogenetically* (i.e., across cultural transmission) by an amplification effect.

References

Birkholz, P., Jackèl, D., & Kroger, B. J. (2006). Construction and control of a three-dimensional vocal tract model. In *Proceedings of the 2006 ieee international conference on acoustics, speech and signal processing* (Vol. 1).

Boë, L.-J., Badin, P., Ménard, L., Captier, G., Davis, B., MacNeilage, P., Sawallis, T. R., & Schwartz, J.-L. (2013). Anatomy and control of the developing human vocal tract: A response to lieberman. *Journal of Phonetics*, *41*(5), 379–392.

Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, *30*(3), 465–484.

Brunner, J., Fuchs, S., & Perrier, P. (2005). The influence of the palate shape on articulatory token-to-token variability. *ZAS Papers in Linguistics*, *4*, 43-67.

Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. *The Journal of the Acoustical Society of America*, *125*(6), 3936–3949.

De Boer, B. (2010). Modelling vocal anatomy's significant effect on speech. *Journal of Evolutionary Psychology*, *8*(4), 351–366.

Dediu, D., Janssen, R., & Moisik, S. R. (2017). Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Language & Communication*, *54*, 9–20.

Fant, G. (1971). *Acoustic theory of speech production: with calculations based on x-ray studies of russian articulations* (Vol. 2). Walter de Gruyter.

Fitch, W. T., Boer, B. de, Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science advances*, *2*(12), e1600723.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of communication disorders*, *39*(5), 350–365.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language* (pp. 121–147). Springer.

Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, *51*(9), 793–809.

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics*, *40*(4), 608–622.

Mooshammer, C., Perrier, P., Geng, C., & Pape, D. (2004). An emma and epg study on token-to-token variability. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, *36*, 47–63.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189–216). Springer.

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the 35th annual meeting of the cognitive science society (cogsci 2013)* (pp. 1348–1353).

Stevens, K. N. (1968). *The quantal nature of speech: Evidence from articulatory-acoustic data.*

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, *17*(1), 3–45.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19.

Tiede, M., Boyce, S. E., Espy-Wilson, C. Y., & Gracco, V. L. (20010). Variability of north american english/r/production in response to palatal perturbation.

Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America*, *115*(5), 2633–2634.

Weirich, M., & Fuchs, S. (2011). Vocal tract morphology can influence speaker specific realisations of phonemic contrasts. In *Proceedings of the international seminar on speech production* (pp. 251–259).

Zhou, X., Espy-Wilson, C. Y., Tiede, M., & Boyce, S. (2007). An articulatory and acoustic study of "retroflex" and "bunched" American English rhotic sound based on MRI. In *Interspeech* (pp. 54–57).