SPEECH-LIKE MOVEMENTS EMERGE FROM SIMULATED PERIORAL MUSCLE ACTIVATION SPACE WITHOUT NEURAL CONTROL

Jonathan de Vries [*1], Ian Stavness [†2], Sid Fels [‡1] and Bryan Gick [♦3]
[1] Interdisciplinary Studies Graduate Program, University of British Columbia, Vancouver, British Columbia, Canada
[2] Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada,
[3] Department of Linguistics, Totem Field Studios, Vancouver, British Columbia, Canada

1 Introduction

Models of speech production to date have lacked realistic and comprehensive properties of physical bodies (see [1]), leaving open the possibility that some mechanisms hitherto ascribed neural control may be reinterpreted as being due to the properties of the biomechanics of the human vocal tract [2, 3]. For example, while neuromuscular modules (muscle synergies) [4] have been proposed as a solution to the degrees of freedom problem in speech biomechanics [5], experimental approaches involving the extraction of muscle synergies are theoretically unable to determine whether such synergies are of neural origin or simply reflect the lower dimensionality of an under-sampled biomechanical/neural task space [6].

As a proof of concept to test the extent to which vocal tract biomechanics may determine speech and expressive facial movements, we created a simplified version of the perioral region using FEM modeling in a physics-based simulator [7]. Systematic simulations using this model enable us to sample the full kinematic/biomechanical space [8, 9] in the absence of central neural control. We aim to use this model to test whether emotive and speech-like movements emerge as self-organizing structures (muscle synergies) in the absence of a direct neural controller.

2 Method

2.1 Model design

We developed a Perioral Simplified Model (POSM) using ArtiSynth (artisynth.org) with significantly reduced degrees of freedom compared to the full face models in Artisynth (see, e.g., [10]). Simplifying the model increased model robustness and stability, reduced computation time, increased model clarity, and allowed analysis techniques to be developed for future use in more complex models.

A finite elemental model (FEM) torus was fitted to the BadinFemMuscleFaceDemo in ArtiSynth to create a simplified biomechanical model of the perioral region (Figure 1). This allows a volumetric, rather than "deformed skin", simulation to be used, which deforms according to material properties of human muscle, skin, and fascia. Five muscle groups were modeled: Marginal and peripheral orbicularis oris (*OOM* and *OOP*, respectively) models use transversely-isotropic FEM-model muscle material and contract in a sphincter-like manner; the *OOP* incorporates separate models for medial and distal concentricities (*OOPm* and *OOPd*) to balance model simplicity with anatomical observation [11]. In addition, two series of point-to-point muscles represent left and right zygomaticus complexes (*ZYGM-l* and *ZYGM-r*, respectively). To capture the interdigitation of zygomaticus major/minor with the *OO* complex, we have modeled both *ZYGM* models with origins at the zygomatic bone and insertions at the mid-lateral FEM blocks containing the *OOP* muscle materials. The musculature for the POSM was selected to allow focus on the more reliably present and robust perioral muscles associated with speech and expression, based on the known wide range of perioral postures produced by the various layers and concentricities of *OO* [11] and the high variation and frequent absence of facial muscles such as risorius [12].

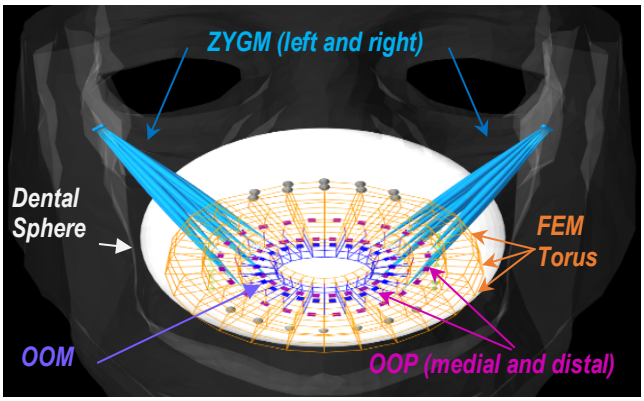

Figure 1: Detail of the perioral simplified model.

The skeletal structures for the POSM (mandible, maxilla, and dentition) are approximated using a smooth, rigid body ellipsoid, with four FEM block nodes fixed at the nasal bone. In addition to the skeletal surface yielding stable results during contact, this surface proved crucial as a backstop against which the model was able to deform, without which the speech and expressive postures were not possible.

2.2 Simulation methodology

All combinations of the five above muscles groups at four levels of activation (0, 0.1, 0.2, 0.3 of maximum excitation as values beyond these resulted in no posture change) were simulated sequentially using adapted Python scripts. Activation sequences were started at zero, held for 200 ms before activation, and achieved target activation after 1000 ms. Once model equilibrium was reached, the resulting posture for that simulation was saved using 3D coordinates for each FEM node as a flattened vector. Aggregation of these posture vectors resulted in a 15504 by 3456 matrix

[*] devries@alumni.ubc.ca
[†] ian.stavness@usask.ca
[‡] ssfels@ece.ubc.ca
[♦] gick@mail.ubc.ca

representing a discretized sampling of the biomechanical state space of our model.

2.3 Analysis methodology

Data were analyzed using t-distributed stochastic neighbourhood embedding (t-SNE) to show regions of the biomechanical space (see [13]) with post-hoc qualitative assessment of kinematic outputs (contact authors for t-SNE and ArtiSynth user-defined parameters). Analyses were repeated 20 times to minimize the objective function and construct a representative visualization of the data. Clusters of speech-like movements were visually obtained by comparing the activation of each muscle across individual simulations. Areas in the t-SNE visualization where one or two muscles show distributional constraint were corroborated in ArtiSynth to determine their qualitative similarity to speech-like or emotive expressions.

3 Results

Figure 2 plots the outputs of the simulations using t-SNE to cluster kinematically similar outputs. Qualitative evaluation revealed that the circled clusters in Figure 1 correspond roughly to lip spreading (top circled cluster), lip protrusion (middle circled cluster), and lip closure (bottom circled cluster).

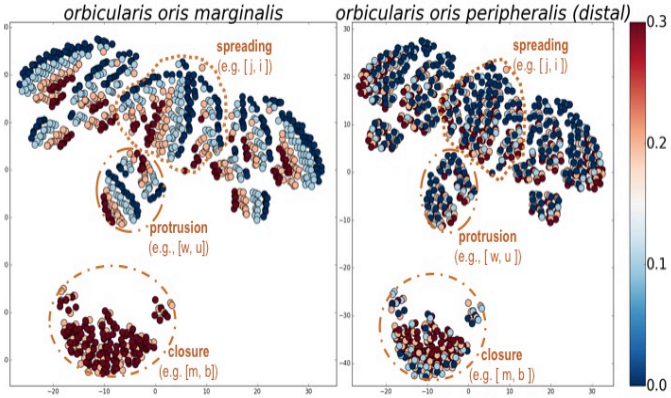

Figure 2 t-SNE (perplexity = 102) visualizations of the POSM state space (each data point corresponds to the output of a complete individual simulation). Left and right visualizations are coloured by muscle activation, here highlighting *OOM* and *OOPd*, respectively. All non-circled data points indicate simulations that display asymmetric postures due to increased force of one *ZYGM* exciter over another.

4 Discussion and conclusion

Systematic simulations enabled us to sample the full kinematic/biomechanical space. t-SNE visualizations of the resulting biomechanical state space show that clusters of similar movements emerge from this space without a direct neural controller, i.e., a uniform sampling of the activation space elicits non-uniform kinematic outputs. These clusters, moreover, fall into ecologically relevant qualitative categories (lip spreading, protrusion and closure), despite competing activation from surrounding muscles.

These results suggest that biomechanics may play a primary role in the near-universal emergence of specific movement categories in emotive and speech-related movements. Because we see similar postures appearing across human populations in speech and facial expression, our analysis suggests that the emergence of these postures may be due at least in part to robustness of these postures to activation noise from surrounding muscles.

The combination of numerical (ArtiSynth and t-SNE) and qualitative procedures used in this study allows us to determine whether similar postures that emerge from the kinematic and biomechanical state space in the absence of neural control resemble speech/emotive-like expressions, providing a proof-of-concept demonstration of this combination of methods as a way of capturing emergent output categories from biomechanical simulations.

Acknowledgments

We acknowledge assistance of L. Ward and G. Truong with analysis; P. Anderson, A. Sanchez and J. Lloyd with ArtiSynth; K. Radford with formatting. Funding from NSERC & NIH Grant DC-002717 to Haskins Laboratories.

References

[1] Gick, B., Schellenberg, M., Stavness, I., & Taylor, R. (2018). Articulatory Phonetics. In W. F. Katz & P. Assmann (eds.). *The Routledge Handbook of Phonetics*. Ch 5. NY: Taylor & Francis.

[2] Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement.* Hillsdale, N.J: L. Erlbaum Assoc.

[3] Stevens, K. (1989). On the quantal nature of speech. *J. Phon.*, 17(1-2): 3-45.

[4] Bizzi, E., & Cheung, V. (2013). The neural origin of muscle synergies. *Front. Compu. Neurosci.*, 7, 51.

[5] Gick, B., & Stavness, I. (2013). Modularizing speech. *Front. Psych.*, 4.

[6] Kutch, J., & Valero-Cuevas, F. (2012). Challenges and new approaches to proving the existence of muscle synergies of neural origin. *PLoS Compu. Biol.*, 8(5)

[7] Lloyd, J. E., Stavness, I., & Fels, S. (2012). ArtiSynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. Berlin: Springer. P. 355-394.

[8] Kutch, J. J., & Valero-Cuevas, F. J. (2011). Muscle redundancy does not imply robustness to muscle dysfunction. *J.Biomech.*, 44(7), 1264-1270.

[9] Gick, B., Allen, B., Roewer-Despres, F. & Stavness, I. (2017). Speaking tongues are actively braced. *JSLHR* 60(3), 494-506.

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

[11] Stavness, I., Nazari, M.A., Perrier, P., Demolin, D. & Payan, Y. (2013). A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. *JSLHR* 56(3): 878-90.

[12] Waller, B. M., Cray, J. J., Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion* 8(3): 435–9.

[13] van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9, 2579-2605.