A Biomechanical Model For Infant Speech And Aerodigestive Movements

Connor Mayer [*1]**, Ian Stavness** [†2] **and Bryan Gick** [‡3, 4]

[1] Department of Linguistics, University of California, Los Angeles, California, USA
[2] Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[3] Department of Linguistics, University of British Columbia, Vancouver, British Columbia, Canada
[4] Haskins Laboratories, New Haven, Connecticut, USA

1 Introduction

Speech is arguably the most complex voluntary movement behavior in the natural world. A central question in speech research is how infants are able to learn speech movements so rapidly and with such limited input.

Computational modeling has become an increasingly promising method for understanding adult speech. Biomechanical modeling allows researchers to study realistic 3D models of the vocal tract built from anatomical and physiological data from perspectives that are difficult to achieve otherwise. Biomechanical modeling of the infant vocal tract will enable similar research into questions relating to speech acquisition.

For example, one appealing proposal regarding speech acquisition is that some core speech movements may build on phylogenetically-encoded aerodigestive actions such as breathing, swallowing and suckling [1, 2]. The idea that neuromuscular primitives such as those relating to digestion are used to bootstrap more complex movement has been suggested in other domains [3, 4], and proposing it as a mechanism in speech acquisition is in accord with neurological [5], clinical [6], and kinematic [7] evidence relating speech and aerodigestive movements.

This hypothesis has proven difficult to test, in part because the anatomy of the vocal tract makes most standard methods such as EMG too invasive or imprecise. The development of biomechanical models of the vocal tract has recently allowed researchers to evaluate some of the predictions made by this hypothesis, showing, for example, that swallowing and movements used in speech occupy similar areas in muscle activation space [8]. A shortcoming of this research is that it has used models of an adult vocal tract to test proposals about infant speech. Adult and infant vocal tracts differ not only in their size, but also in the relative proportions of their structures [9, 10].

In this paper we present a preliminary model of an infant vocal tract using the 3D biomechanical simulation platform Artisynth (e.g. [11, 12]). This model is generated by modifying an existing vocal tract model implemented in Artisynth to conform to the proportions of a mid-sagittal CT image of an infant vocal tract. The model will be capable of simulating both swallowing and simple speech movements, and the results of simulations using this model will provide useful insight into infant motor control, supplementing

[*] connormayer@ucla.edu
[†] ian.stavness@usask.ca
[‡] gick@mail.ubc.ca

evidence from other domains.

2 Methods

Artisynth contains a model of an adult vocal tract called Frank [13]. We used Frank as a starting point, modifying its relative proportions to match more closely those of an infant. To determine the proportions, we compared a mid-sagittal CT image of an 11-month-old male infant vocal tract to a mid-sagittal image of the Frank vocal tract. A set of twelve biometric measurements were modified from previous studies of infant vocal tract development [9, 10].

These measurements were (1) *vocal tract length*: the curvilinear distance along the midline of the vocal tract from the superior point of the larynx to a line tangential to the incisors; (2) *hard palate length*: the curvilinear distance along the hard palate contour from the anterior base of the incisor to the beginning of the soft palate; (3) *soft palate length*: the curvilinear distance along the soft palate contour from the posterior edge of the hard palate to the inferior edge of the uvula; (4) *mandibular length*: the length of a straight line from the mental protuberance to the anterior point of the thyroid cartilage; (5) *tongue length*: the curvilinear distance along the tongue surface from the vallecula to the tongue tip; (6) *hyoid height*: the vertical distance from the posterior nasal spine to the superior edge of the hyoid bone; (7) *larynx height*: the vertical distance from the posterior nasal spine to the superior edge of the vocal folds; (8) *oropharynx length*: the vertical distance from the inferior edge of the uvula to the superior edge of the vocal folds; (9) *vocal tract horizontal length*: the horizontal distance from a line tangent to the incisors to the back of the nasopharynx; (10) *anterior cavity length*: the horizontal distance from a line tangent to the incisors to the intersection with a vertical line from the anterior edge of the glottis to the palate; (11) *oropharynx width*: the horizontal distance from the back of the nasopharynx to the intersection with a vertical line from the anterior edge of the glottis to the palate; and (12) *mandible height*: the length of a straight line from the inferior point of the mandible to the superior point of the lower incisors.

Because the images at are different scales, all measurements were normalized by dividing each distance by the mandible height. The infant and Frank vocal tracts with measurements overlaid are shown in Figure 1.

The most substantial difference between the infant and Frank vocal tracts is that the larynx and hyoid are higher in the infant. This is reflected in the *vocal tract length*, *hyoid level*, *larynx level*, and *oropharynx length* measurements.
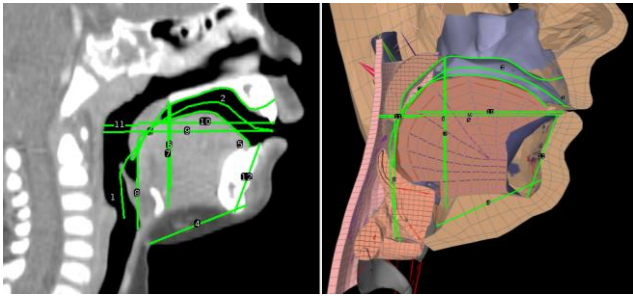
Figure 1: The infant vocal tract (left) and Frank vocal tract model (right) with overlaid measures.

In order to quantify similarity between the infant vocal tract and the Artisynth models, we designed a measure θ similar to the sum of squared errors:

$$\theta = \sum_{i=1}^{m} I_i \left(M_i - I_i\right)^2$$

Where i indexes over the set of measurements, M_i is the normalized measurement from the model, and I_i is the normalized measurement from the infant CT image. This measure penalizes differences in larger measurements more than differences in shorter measurements, and lower values indicate a closer correspondence to the infant vocal tract.

The structures in the Frank vocal tract were manually modified using rigid body translations and affine transformations to bring them in line with the proportions of the infant vocal tract. We also removed excess tissue, including the face. The same measurements were then carried out on the infant vocal tract model and compared to those of the original Frank model (Figure 2).
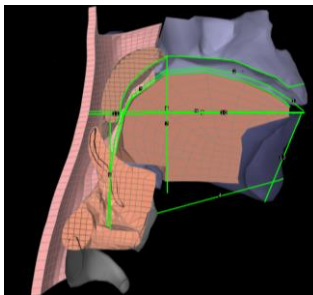

Figure 2: Infant vocal tract model with overlaid measures.

3   Results

For the original Frank model, $\theta = 1.438$, while the infant vocal tract model, $\theta = 0.373$. This indicates that the infant vocal tract model in Artisynth corresponds more closely to the dimensions of the infant vocal tract than the original Frank model.

4   Discussion and future work

The infant vocal tract model presented in this paper will allow researchers to simulate both aerodigestive and speech movements. The results of these simulations will help to supplement neurological, clinical, and kinematic evidence bearing on hypotheses relating to the acquisition of speech and aerodigestive movements. In addition, we plan to further increase the accuracy of the infant model by validating the muscle insertions and replacing the maxilla and mandible with 3D models generated from infant data.

References

[1] MacNeilage, P. (2008). *The Origin of Speech*. Oxford and New York: Oxford University Press.

[2] Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. Christiansen & S. Kirby (Eds.), *Language Evolution*. Oxford and New York: Oxford University Press.

[3] Wolpert, D.M., Ghahramani, Z., & Flanagan, J.R. (2001). Perspectives and problems in motor learning. *TRENDS in Cognitive Science, 5*(11), 487-94.

[4] Dominici, N., Ivanenko, Y.P., Cappellini, G., d'Avella, A., Mondi, V., Cicchese, M., Fabiano, A., Silei, T., Di Paolo, A., Giannini, C., Poppele, R.E., & Lacquaniti, F. (2011). Locomotor Primitives in Newborn Babies and Their Development. *Science*, *334*(6058), 997-9.

[5] Martin, R.E., MacIntosh, B.J., Smith, R.C., Barr, A.M., Stevens, T.K., Gati, J.S., & Menon, R.S. (2004). Cerebral areas processing swallowing and tongue movement are overlapping but distinct: a functional magnetic resonance imaging study. *J. Neurophysiology, 92*(4), 2428-43.

[6] LaGorio, L.A., Carnaby-Mann, G.D., Crary, M.A. (2008). Cross-system effects of dysphagia treatment on dysphonia: a case report. *Cases Journal, 1*(1):1-67.

[7] Green, J.R., Moore, C.A., Higashikawa, M., & Steeve, R.W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *J. of Speech, Language & Hearing Research, 43*(1), 239-55.

[8] Mayer, C., Roewer-Despres, F., Stavness, I., & Gick, B. (2017). Do innate stereotypies serve as a basis for swallowing and learned speech movements? *Behavior and Brain Sciences, 40*.

[9] Vorperian, H.K., Kent, R.D., Lindstrom, M.J., Kalina, C.M., Gentry, L.R., & Yandell, B.S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *J. Acous. Soc. Am., 117*(1), 338-50.

[10] Vorperian, H.K., Wang, S., Chung, M.K., Schimek, E.M., Durtschi, R.B., Kent, R.D., Ziegert, A.J., & Gentry, L.R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Acous. Soc. Am, 125*(3), 1666-78.

[11] Stavness, I., Lloyd, J.E., & Fels, S.S (2012). Automatic Prediction of Tongue Muscle Activations Using a Finite Element Model. *J. Biomechanics, 45*(16), 2841-8.

[12] Gick, B., Anderson, P., Chen, H., Chiu, C., Kwon, H.B., Stavness, I., Tsou, L., & Fels, S. (2014). Speech function of the oropharyngeal isthmus: A modeling study. *Computer Methods in Biomechanics & Biomedical Engineering: Imaging & Visualization, 2*(4), 217-22.

[13] Anderson, P., Fels, S., Harandi, N.M., Ho, A., Moisik, S., Sanchez, C.A., Stavness, I., Tang, K. (2017). FRANK: a hybrid 2D biomechanical model of the head and neck. In Y. Payan & J. Ohayon (Eds.), *Biomechanics of Living Organisms*. Academic Press, 413-47.