

## Understanding how we speak using computational models of the vocal tract

Connor Mayer - [connormayer@ucla.edu](mailto:connormayer@ucla.edu)

Department of Linguistics - University of California, Los Angeles

Ian Stavness - [ian.stavness@usask.ca](mailto:ian.stavness@usask.ca)

Department of Computer Science - University of Saskatchewan

Bryan Gick - [gick@mail.ubc.ca](mailto:gick@mail.ubc.ca)

Department of Linguistics - University of British Columbia; Haskins Labs

Popular version of poster 5aSC1, “A biomechanical model for infant speech and aerodigestive movements”

Presented Friday morning, November 9, 2018, 8:30-11:30 AM, Upper Pavilion

176th ASA Meeting and 2018 Acoustics Week in Canada, Victoria, Canada

Speaking is arguably the most complex voluntary movement behaviour in the natural world. Speech is also uniquely human, making it an extremely recent innovation in evolutionary history. How did our species develop such a complex and precise system of movements in so little time? And how can human infants learn to speak long before they can tie their shoes, and with no formal training?

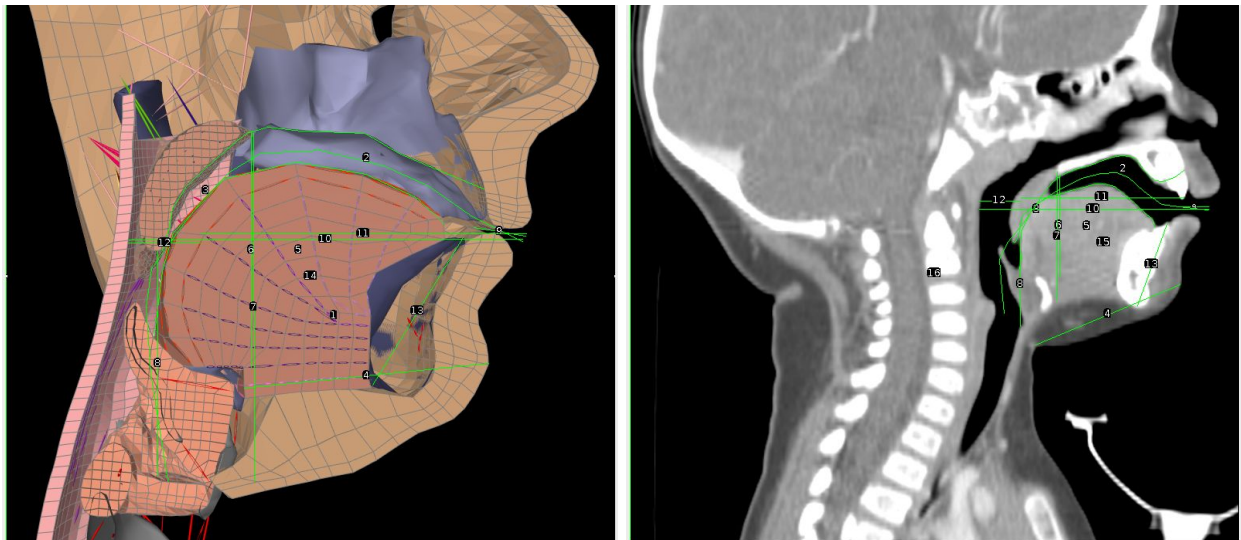
Answering these questions requires a deep understanding of how the human body makes speech sounds. Researchers have used a variety of techniques to understand the movements we make with our vocal tracts while we speak – acoustic analysis, ultrasound, brain imaging, and so on. While these approaches have increased our understanding of speech movements, they are limited. For example, the anatomy of the vocal tract is quite complex, and tools that measure muscle activation, such as EMG, are too invasive or imprecise to be used effectively for speech movements.

Computational modeling has become an increasingly promising method for understanding speech. The biomechanical modeling platform Artisynt ( <https://www.artisynt.org> ), for example, allows scientists to study realistic 3D models of the vocal tract that are built using anatomical and physiological data.

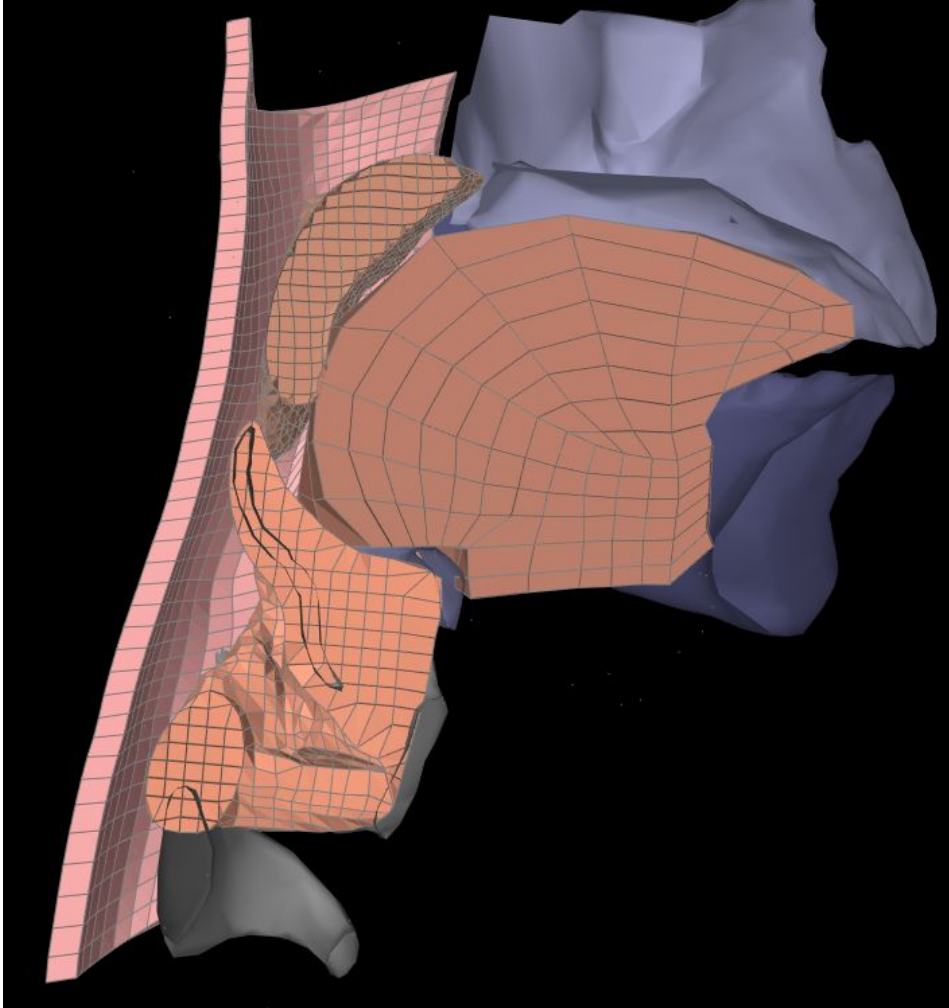
These models can be used to see aspects of speech that are hard to visualize using other tools. For example, we can see what shape the tongue takes when a specific set of muscles activates. Or we can have the model perform a certain action and measure aspects of the outcome, like having the model produce the syllable “ba” and looking at how much the lips deform by mutual compression during their contact in the /b/ sound. We can also predict how changes to typical vocal tract anatomy, such as the removal of part of the tongue in response to oral cancer, affect the ability to perform speech movements.

In our project at the 176th ASA Meeting, we present a model of the vocal tract of an 11 month old infant. A detailed model of the adult vocal tract named 'Frank' has already been implemented in Artisynt, but the infant vocal tract has different proportions than an adult vocal tract. Using Frank as a starting point, we modified the relative scale of the different structures based on measurements taken from CT scan images of an infant vocal tract (see Figure 1).

Going forward, we plan to use this infant vocal tract model (see Figure 2) to simulate both aerodigestive movements and speech movements. One of the hypotheses for how infants learn to speak so quickly is that they build on movements they can carry out at birth, such as swallowing or suckling. The results of these simulations will help supplement neurological, clinical, and kinematic evidence bearing on this hypothesis. In addition, the model will be generally useful for researchers interested in the infant vocal tract.



*Figure 1: Left: A cross-section of the Frank model of an adult vocal tract with measurement lines. Right: A cross-sectional CT scan image of an 11 month old infant with measurement lines. The relative proportions of each vocal tract were compared to generate the infant model.*



*Figure 2: A modified Frank vocal tract conforming to infant proportions.*