Looking Behind the Face: Research on Robotic and Computational Modeling Oropharyngeal Anatomy for Speech Synthesis

Sidney Fels
Dept. of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada, V6T 1Z4
ssfels@ece.ubc.ca

In this workshop, I will present work on our 3D articulatory speech synthesizer, called Artisynth, that is based on a framework for modeling anatomical structures. Our research has led to the creation of highly accurate computational 3D articulator models as well as a robotic 6 degree-of-freedom jaw.

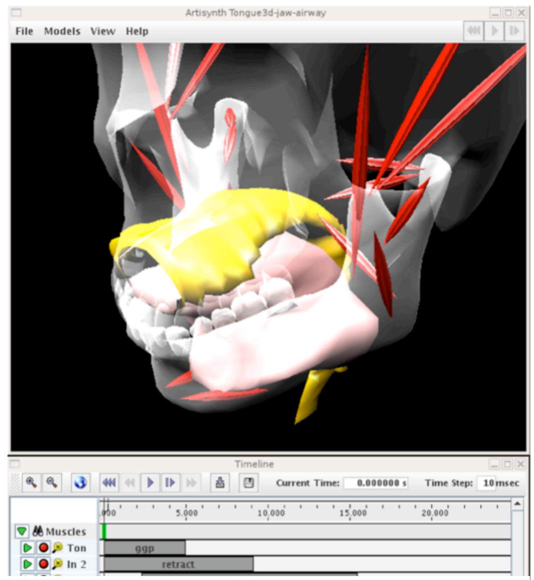

Figure 0: Artisynth 3D Model Viewer with jaw, tongue and airway (top) and Timeline interface (bottom).

ArtiSynth is a 3D biomechanical simulation platform directed toward modeling the vocal tract and upper airway. It provides an open-source environment in which researchers can create and interconnect various kinds of dynamic and parametric models to form a complete integrated biomechanical system which is capable of articulatory speech synthesis. An interactive graphical Timeline (Figure 1) runs the simulation and allows the temporal arrangement of input/output channels to control or observe properties of the model's components. Library support is available for particle-spring and rigid body systems, finite element models, and spline-based curves and surfaces. To date, these have been used to create a dynamic muscle-based model of the jaw, a deformable tongue model, a deformable airway, and a linear acoustics model, which have been connected together to form a complete vocal tract that produces speech and is drivable both by data and by dynamics such as shown in Figure 1.

As part of our explorations of the vocal tract, we have created an anthropomorphic, 6 DOF robotic jaw capable of producing, in real-time, the complex set of motions described by the human jaw during speech or mastication shown in Figure 2. The jaw is designed to fit within a larger robotic human figure such as the head, neck and torso of the 25 DOF Infanoid [1]. The produced mechanical prototype has been designed to accommodate a prosthetic mandible with dentures. The mechanism fits within the skull of the average man; where it would occupy less than 1/3 of the skull cavity. Two TMJs (temporomandibular joints) support the prosthesis, where each is driven by a 3 DOF parallel manipulator. In order to combine the motion of both manipulators each TMJ is capable of 3 DOF. The system is controlled via a USB port using software that models the human skull including collision detection mechanisms. The jaw allows for linear control, zero-backlash, and up to three times exaggerated mobility ranges making it also suitable for speech research, facial gesture affect research and dentistry applications.

We are continuing to improve Artisynth but including more complex models as well as fast FEM techniques to achieve interactive rates for the simulation. We expect to create a complete, 3D average vocal tract using our framework. As well, we are working on a 22 DOF robotic head and neck that includes the 6DOF jaw.

Acknowledgment
This research is funded by: Natural Sciences and Engineering Research Council (NSERC), Canada; Advanced Telecommunication Research (ATR) Laboratory, Japan, and Peter Wall Institute of Advanced Studies (PWIAS), Canada. The Artisynth team includes: Eric Vatikiotis-Bateson, Carol Jaeger, Alan Hannam, John Lloyd, Florian Vogt, Kees van den Doel, Ian Stavness and many other contributors.

References
[1] Hideki Kozima: Infanoid: A Babybot that Explores the Social Environment, K. Dautenhahn et al. (eds.), Socially Intelligent Agents: Creating Relationships with Computers and Robots, Kluwer Academic Publishers, pp.157-164, 2002.

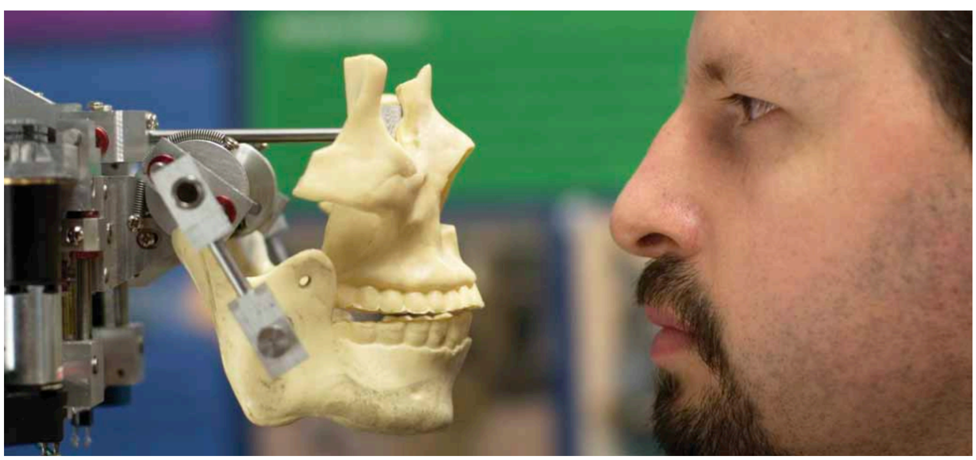

Figure 2: Photo of our 6 DOF robotic jaw and its creator, Edgar Flores.