Developing Physically-Based, Dynamic Vocal Tract Models using ArtiSynth

Sidney Fels[1] **, John E. Lloyd**[1] **, Kees van den Doel**[2]**,**
Florian Vogt[1] **, Ian Stavness**[1] **, Eric Vatikiotis-Bateson**[3]

[1] Department of Electrical and Computer Engineering,

[2]Department of Computer Science,

[3]Department of Linguistics,
University of British Columbia, Canada

ssfels@ece.ubc.ca

Abstract. *We describe the process of using ArtiSynth, a 3D biomechanical simulation platform, to build models of the vocal tract and upper airway which are capable of simulating speech sounds. ArtiSynth allows mass-spring, finite element, and rigid body models of anatomical components (such as the face, jaw, tongue, and pharyngeal wall) to be connected to various acoustical models (including source filter and airflow models) to create an integrated model capable of sound production. The system is implemented in Java, and provides a class API for model creation, along with a graphical interface that permits the editing of models and their properties. Dynamical simulation can be interactively controlled through a "timeline" interface that allows online adjustment of model inputs and logging of selected model outputs. ArtiSynth's modeling capabilities, in combination with its interactive interface, allow for new ways to explore the complexities of articulatory speech synthesis.*

1. Introduction

Computer simulation of anatomical and physiological processes is becoming a popular and fruitful technique in a variety of medical application areas. Use of such techniques has become common for inquiries into articulatory speech synthesis and the understanding of speech production mechanisms. This has been aided by advancements in the computer graphics and animation fields which have spawned a variety of schemes for creating fast and accurate physically-based simulation. In this paper, we describe the most recent version of ArtiSynth, a general purpose biomechanical simulation platform focused toward creating integrated 3D models of the vocal tract and upper airway, including the head, tongue, face, and jaw. New features within the system provide functionality for easily creating, modifying and interacting with complex biomechanical models and simulating speech sounds. ArtiSynth's capabilities have a wide range of applications in medicine, dentistry, linguistics, and speech research. Specific examples include (a) studying the physiological processes involved in human speech production with the goal of creating a 3D articulatory speech synthesizer, (b) planning for maxillo-facial and jaw surgery, and (c) analyzing medical phenomena such as obstructive sleep apnea (OSA).

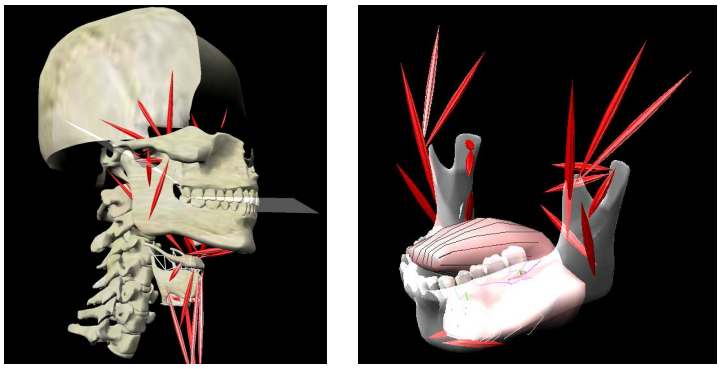

Figure 1. (a) Jaw and laryngeal model implemented using ArtiSynth, (b) Jaw model connected to a model of the tongue

Capabilities which have been added to ArtiSynth since our last report (Fels et al., 2005) include:

- 3D finite element models (section 3);
- Collision detection between rigid bodies and finite element models (section 3);
- An enhanced airway including nasal passages and coupling to a 3D tongue model (section 3.3);
- Aero-acoustic improvements including Rosenberg or waveform excitation, the Ishikawa-Flanagan two mass model, a 1D real-time Navier-Stokes solver, and fricative modeling (section 3.3);
- Editing of model components and their properties, enabled by selection through both the graphical display and a navigation panel (section 4);
- Improved methods for interactive simulation control (section 5).

ArtiSynth supports the construction and integration of a diverse set of anatomical and acoustic models, such as those listed in section 2. Models which we have created to date include a dynamic jaw/laryngeal model (Stavness et al., 2006), a muscle activated finite element model of the tongue (Vogt et al., 2006), and an integrated airway/tongue model capable of synthesizing speech sounds (Doel et al., 2006). Figure 1 shows the jaw/laryngeal model, along with its connection to the tongue, and the airway/tongue model is shown in Figure 3. Demos, information, and software can be obtained from *www.artisynth.org*.

2. Related Work

Many researchers from different areas have developed independent models of vocal tract and airway components, including the tongue, larynx, lips, and face, using both parametric and biomechanical models. Specific attention has been directed towards modeling these structures for predicting sleep apnea (Huang et al., 2005), speech production (Gerard et al., 2006; Dang and Honda, 2004; Sanguineti et al., 1998; Payan and Perrier, 1997), head posturing (Munhall et al., 2004; Koolstra and van Eijden, 2004; Shiller et al., 2001), swallowing (Hiiemae and Palmer, 2003; Palmer et al., 1997; Li et al., 1994; Chang et al., 1998; Berry et al., 1999), and facial movements (Lee et al., 1995; Sifakis et al., 2005; Luboz et al., 2005; Gladilin et al., 2004; Pitermann and Munhall, 2001). The complex aero-acoustical processes that involve the interaction of these anatomical elements with

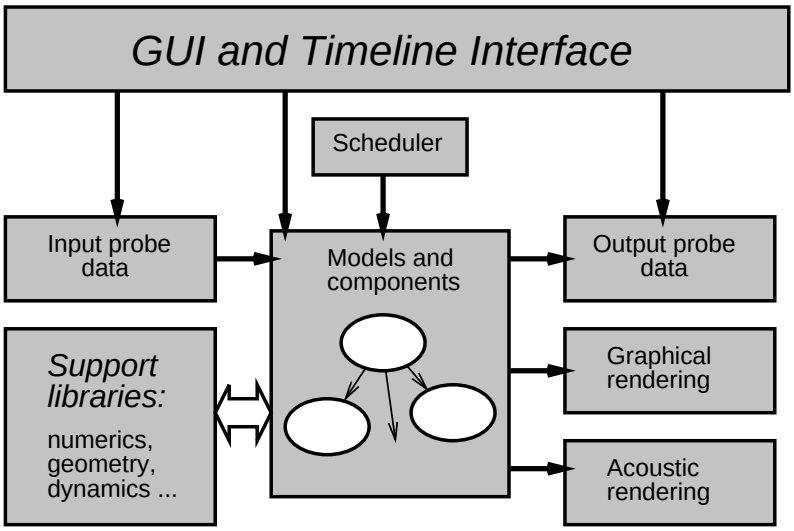

Figure 2. General architecture of the ArtiSynth system.

airflow and pressure waves, and which eventually produce speech, have also been studied (Svancara et al., 2004; Sinder et al., June, 1998).

The geometry and parameters for the above-mentioned anatomical models are usually created by a combination of semi-automatic extraction methods (applied to medical imaging data) and hand tuning procedures (Gerard et al., 2003; McInerney and Terzopolis, 1999; Stavness et al., 2006). Actual simulation is usually performed by either 1) engineering packages such as Femlab, Adams or ANSYS, 2) surgical simulation packages such as SOFA, GiPSi, SimTK or Open Tissue, or 3) movie and computer game software such as Maya, Blender, RealFlow, openODE, or Havok.

3. Modeling Capabilities

3.1. General Framework

The overall architecture of ArtiSynth is shown in Fig. 2. At the center is the system model, which is a hierarchical collection of *models* and their *components*. A *scheduler* is responsible for controlling the dynamic simulation of the models through time, under the control of an external GUI and *Timeline* interface (Sec. 5). Each model implements an `advance` method which is responsible for advancing the model through time.

Models and their components are implemented as Java classes, and provide support for maintaining themselves within a hierarchy, reading and writing to persistent storage, and (optionally) rendering themselves graphically or acoustically. The graphical rendering system is presently implemented using OpenGL, and permits model viewing (through one or more graphical displays), component selection, and various forms of graphical interaction. Model components may also export *properties*, which are attributes with support for setting or getting their values from the ArtiSynth graphical user interface (GUI). Model simulation can be interactively controlled through a *timeline* GUI, as described in Section 5. Creating and editing models is discussed in Section 4.

3.2. Biomechanical Models

The central artifact for creating biomechanical models is a model class called *MechModel*, which implements assemblies of 3D finite element models (FEMs), lineal (point-to-point) springs, particles, and rigid bodies. Bilateral and unilateral constraints are also provided for simulating joints and contacts. MechModel simulates the dynamic evolution of all its components by integrating the dynamics equation

$$\mathbf{M}(\mathbf{x})\,\ddot{\mathbf{x}} \;=\; \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, t) \tag{1}$$

where $\mathbf{M}$, $\mathbf{x}$, and $\mathbf{f}$ denote the aggregate mass matrix, dynamic state, and generalized forces for the entire system. A variety of explicit and implicit integrators are available (e.g., forward Euler, Runga Kutta, backward Euler). Support is also provided for computing and solving the force/state Jacobian $\mathbf{J} = \partial\mathbf{f}/\partial(\mathbf{x}, \dot{\mathbf{x}})$, which is needed by implicit integrators for solving stiff systems (such as those containing FEMs).

FEMs are implemented using the fast linear stiffness-warping approach of (Mueller and Gross, 2004), in which rotation is factored out of the strain tensor in order to reduce distortion. Muscle activation is supported within these FEM models, based on the application of uniform contraction forces along selected element edges, as described in (Vogt et al., 2006).

Specific dynamic components of a MechModel (such as particles or rigid bodies) can be declared non-dynamic, so that their position and velocity are either fixed or controlled by external inputs. This allows parametric control of selected components within a MechModel.

MechModel currently supports collision handling between rigid bodies (using an impulse method similar to that described in (Kaufman et al., 2005)) and between FEMs and inactive rigid bodies (by projecting penetrating FEM nodes onto the rigid body surface). Collision handling is being extended to handle all interactions between rigid and deformable bodies and particles. The collision detection itself is done by computing intersections between the surface meshes of either the FEM or the rigid body, using an oriented bounding box (OBB) hierarchy (Gottschalk et al., 1996).

3.3. Acoustic Modeling

In order to model the aero-acoustical phenomena in the vocal tract we need a model for the airway. In principle, the airway is determined implicitly by its adjacent anatomical components, but as some of these components may not yet be modeled, or may be of limited relevance, we have developed a stand-alone version of the vocal tract airway. Such explicit airway modeling is also described in (Yehia and Tiede, 1997; Honda et al., 2004).

Our airway consists of a mesh-based surface model (Figure 3). The mesh is structured as a sequence of cross-sectional polygons arranged along the airway's length, which permits fast calculation of cross-sectional areas along the airway's center line and hence allows the acoustical chamber to be modeled as a cylindrically symmetric tube. The airway is deformable and changes shape in concert with the anatomical components (such as the tongue) which surround it.

The wave propagation through the vocal tract is modeled using the linearized Navier-Stokes equations which we solve numerically in real-time on a 1D grid using an implicit-explicit Euler scheme (Doel and Ascher, 2006). The method remains stable when small constrictions in the airway generate strong damping. An advantage of this approach over the well-known Kelly-Lochbaum (Kelly and Lochbaum, 1962) tube segment filter model is that the airway can be stretched continuously (when pursing the lips for example) which is not possible with the classical Kelly-Lochbaum model (also available in ArtiSynth) which requires a fixed grid size. The vocal chords are modeled using the

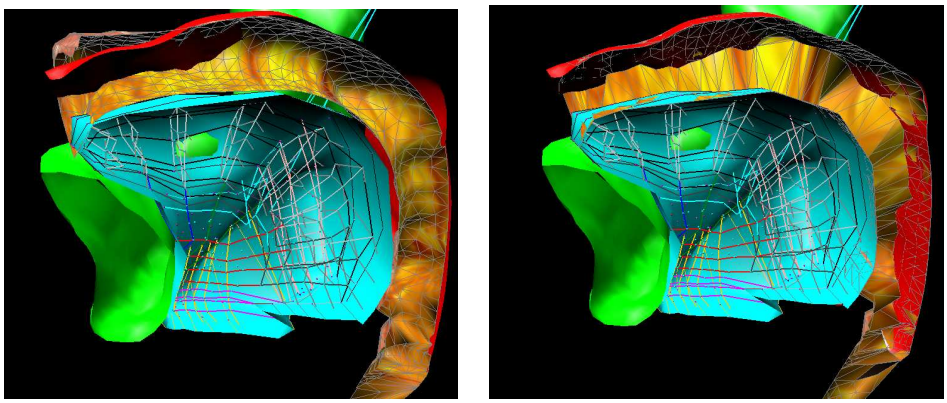

Figure 3. The Airway model and the tongue, palate and jaw meshes before (left) and after registration.

Ishizaka-Flanagan two-mass model (Ishizaka and Flanagan, 1972; Sondhi and Schroeter, 1987). This model computes pressure oscillations as well as glottal flow and is dynamically driven by lung pressure and tension parameters in the vocal chords. The vocal chord model is coupled to the discretized acoustics equation in the vocal tract. Noise is injected at the narrowest constriction and at the glottis according to the model described in (Sondhi and Schroeter, 1987). The resulting model is capable of producing vowels as well as fricatives. Artisynth also supports Rosenberg excitation or sampled waveforms.

The airway exerts forces on the anatomical structures that are connected to it. The air velocity u along the airway determines the pressure P through the steady state solution of the Navier-Stokes equation which is equivalent to Bernoulli's law.

4. Model Creation and Editing

Models may be created either directly in Java, using appropriate method calls, or by constructing a text file description which conforms to a standard ArtiSynth file format. This file format also facilitates saving and restoring models which have been interactively edited.

We have implemented some support for reading in model structures described using other application formats. This includes the Alias Wavefront `.obj` format for describing meshes, the ANSYS file format for tetrahedral and hexahedral FEMs, and landmark data produced by Amira. More generally, the creation of ArtiSynth models often involves extracting model geometry from medical image data (using tools such and

Rhino and Amira) and then using this geometry to define deformable or rigid dynamic structures. An overview of this process is given in Stavness et al. (2006).

It is also possible to interactively edit models with the ArtiSynth GUI. Currently, using navigation and selection tools, researchers may set property values, transform geometry (translation, rotation, and scaling), and add or delete components.

5. Interactive Simulation

ArtiSynth provides means to "instrument" a simulation by attaching input and output *probes* to the models or their components. Input probes are data streams which can set control inputs or modulate parameter values over time. For example, they may supply time-varying sets of activation levels to control a muscle model, or primary vocal cord waveforms to drive an aero-acoustical model. Output probes are data streams which can record model variables or properties for a portion of the simulation. For example, they may be used to log the motions of an anatomical component such as the jaw, or the final waveform produced by an aero-acoustical model.

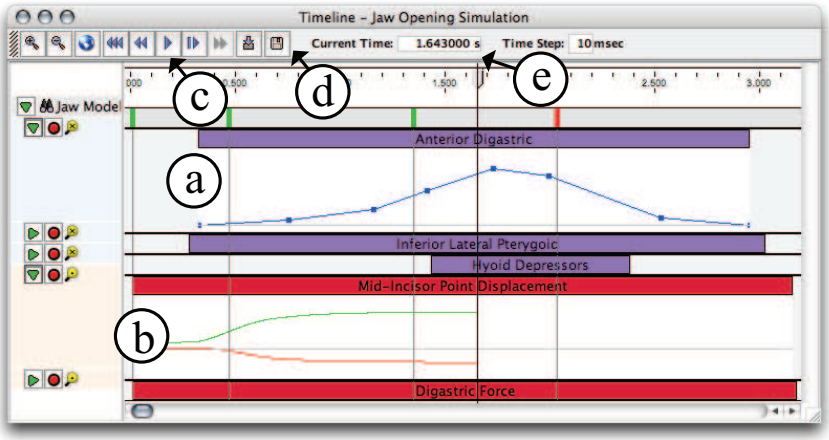

Figure 4. The ArtiSynth Timeline controlling a jaw motion simulation: (a) input probe (e.g. muscle activation), (b) output probe (e.g. incisor position), (c) play controls, (d) save / load buttons, (e) current time cursor.

Input/output probes can be scheduled graphically by arranging icons on a *Timeline* GUI component, which also provides play controls for running, pausing, and stepping the simulation. Fig. 4 shows a full screen shot of the ArtiSynth GUI, with control inputs arranged on the Timeline. Intuitively, our approach is to blend the timeline features of movie edit applications with the simulation environment of 3D modeling tools.

Attaching probes to a model can be done interactively, using a *probe edit* window. This allows a user to attach input or output probes to model component properties which are selected using the navigation and selection mechanism mentioned in Section 4.

A particular model configuration, with probe settings, can be saved and restored as a *workspace*, facilitating incremental development as well as sharing of work. Most of the ArtiSynth GUI actions are associated with well-defined API calls supporting direct control of the simulations with Java scripts. These can be specified either through a Jython console or through the Matlab java interface.

6. Summary

We have reported on the ongoing development of ArtiSynth, an open source environment for combining biomechanical and acoustical models toward creating complete 3D simulations of the vocal tract and upper airway. The new features reported here have been inspired by using ArtiSynth to create new, complex models of the jaw, tongue and combination of the two for speech synthesis. We will continue to develop ArtiSynth, and we invite other researchers to either use the system or contribute to the project.

Acknowledgments

This work was supported by NSERC, the Peter Wall Institute for Advanced Studies and the Advanced Telecommunications Research Laboratory (Japan). We also gratefully acknowledge the many contributions of the people and organizations listed on the ArtiSynth website (www.artisynth.org).

References

Berry, D. A., Moon, J. B., and Kuehn, D. P. A finite element model of the soft palate. *Cleft Palate-Craniofacial J*, 36(3):217–223, 1999.

Chang, M. W., Rosendall, B., and Finlayson, B. A. Mathematical modeling of normal pharyngeal bolus transport: A preliminary study. *J of Rehab Res and Dev*, 35(3):327–334, 1998.

Dang, J. and Honda, K. Construction and control of a physiological articulatory model. *JASA*, 115(2):853–870, 2004.

Doel, K. v. d. and Ascher, U. Staggered grid discretization for the Webster equation. *in preparation*, 2006.

Doel, K. v. d., Vogt, F., English, E., and Fels, S. Towards Articulatory Speech Synthesis with a Dynamic 3D Finite Element Tongue Model. *submitted to 7th ISSP*, 2006.

Fels, S., Vogt, F., van den Doel, K., Lloyd, J. E., and Guenther, O. Artisynth: Towards realizing an extensible, portable 3D articulatory speech synthesizer. In *International Workshop on Auditory Visual Speech Processing*, pages 119–124, 2005.

Gerard, J. M., Perrier, P., and Payan, Y. *3D biomechanical tongue modelling to study speech production*, pages 85–102. Psychology Press: New York ,USA, 2006.

Gerard, J. M., Wilhelms-Tricarico, R., Perrier, P., and Payan, Y. A 3D dynamical biomechanical tongue model to study speech motor control. *Rec Res Dev in Biomech*, 1:49–64, 2003.

Gladilin, E., Zachow, S., Deuflhard, P., and Hege., H. C. Anatomy and physics based facial animation for craniofacial surgery simulations. *Med & Bio Eng & Comp*, 42(2):167–170, 2004.

Gottschalk, S., Lin, M. C., and Manocha, D. OBBtree: A hierarchical structure for rapid interference detection. *ACM Trans on Graphics*, 15(3), 1996.

Hiiemae, K. and Palmer, J. Tongue movements in feeding and speech. *Crit Rev Oral Biol Med*, 14(6):413–29, 2003.

Honda, K., Takemoto, H., Kitamura, T., and Fujita, S. Exploring human speech production mechanisms by MRI. *IEICE Info Sys*, E87-D:1050–1058, 2004.

Huang, Y., White, D., and Malhotra, A. The impact of anatomical manipulations on pharyngeal collapse: Results from a comp. model of the normal human airway. *Chest*, 128:1324, 2005.

Ishizaka, K. and Flanagan, J. L. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Systems Technical Journal*, 51:1233–1268, 1972.

Kaufman, D. M., Edmunds, T., and Pai, D. K. Fast frictional dynamics for rigid bodies. *ACM Trans. Graph.*, 24(3):946–956, 2005. ISSN 0730-0301.

Kelly, K. L. and Lochbaum, C. C. Speech Synthesis. In *Proc. Fourth ICA*, 1962.

Koolstra, J. H. and van Eijden, T. M. G. J. Functional significance of the coupling between head and jaw movements. *J Biomech*, 37(9):1387–1392, 2004.

Lee, Y., Terzopoulos, D., and Waters, K. Realistic modeling for facial animation. In *SIGGRAPH*, pages 55–62, 1995.

Li, M., Brasseur, J., and Dodds, W. Analyses of normal and abnormal esophageal transport using computer simulations. *Am J Physiol*, 266(4.1):G525–43, 1994.

Luboz, V., Chabanas, M., Swider, P., and Payan, Y. Orbital and maxillofacial computer aided surgery: patient-specific finite element models to predict surgical outcomes. *Comput Methods Biomech Biomed Engin*, 8(2):259–65, 2005.

McInerney, T. and Terozopolis, D. Topology adaptive deformable surfaces for medical image volume segmentation. *IEEE Trans on Med Im*, 18(10):840–850, 1999.

Mueller, M. and Gross, M. Interactive virtual materials. In *GI*, pages 239–246, 2004.

Munhall, K., Jones, J., Callan, D., Kuratate, T., and Vatikiotis-Bateson, E. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psych Science*, 15:133–137, 2004.

Palmer, J., Hiiemae, K., and Lui, J. Tongue-jaw linkages in human feeding. *Arch Oral Biol*, 42: 429–441, 1997.

Payan, Y. and Perrier, P. Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis. *SC*, 22(2):185–205, 1997.

Pitermann, M. and Munhall, K. An inverse dynamics approach to face animation. *JASA*, 110: 1570–1580, 2001.

Sanguineti, V., Laboissiere, R., and Ostry, D. J. A dynamic biomechanical model for neural control of speech production. *JASA*, 103:1615–1627, 1998.

Shiller, D., Ostry, D., Gribble, P., and Laboissiere, R. Compensation for the effects of head acceleration on jaw movement in speech. *J Neurosci*, 21:6447–6456, 2001.

Sifakis, E., Neverov, I., and Fedkiw, R. Automatic determination of facial muscle activations from sparse motion capture marker data. In *ACM SIGGRAPH*, 2005.

Sinder, D. J., Krane, M. H., and Flanagan, J. L. Synthesis of fricative sounds using tan aeroacoustic noise generation model. In *Proc. ASA Meet.*, June, 1998.

Sondhi, M. M. and Schroeter, J. A Hybrid Time-Frequency Domain Articulatory Speech Synthesizer. *IEEE Trans on Acoustics, Speech, and Signal Processing*, ASSP-35(7):955–967, 1987.

Stavness, I., Hannam, A. G., Lloyd, J. E., and Fels, S. An integrated, dynamic jaw and laryngeal model constructed from ct data. *Proc ISBMS06 in Springer LNCS 4072*, pages 169–177, 2006.

Svancara, P., Horacek, J., and Pesek, L. Numerical modeling of production of Czech vowel /a/ based on FE model of vocal tract. In *Proc ICVPB*, 2004.

Vogt, F., Lloyd, J. E., Buchaillard, S., Perrier, P., Chabanas, M., Payan, Y., and Fels, S. S. Investigation of efficient 3D finite element modeling of a muscle-activated tongue. *Proceedings of ISBMS 06 in Springer LNCS 4072*, pages 19–28, 2006.

Yehia, H. C. and Tiede, M. A parametric three-dimensional model of the vocal-tract based on MRI data. In *Proc ICASSP*, pages 1619–1625, 1997.