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).

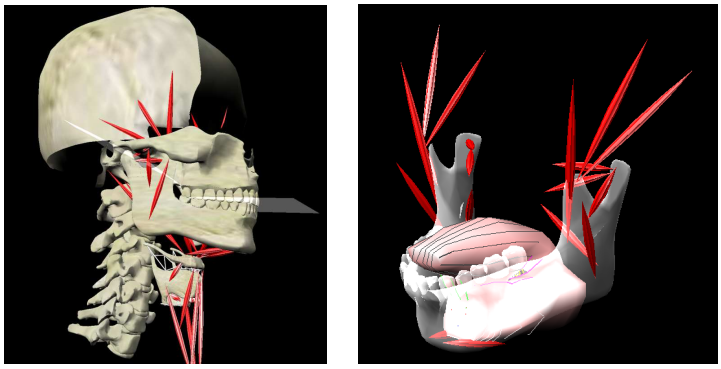

Figure 1. (a) Jaw and laryngeal model implemented using ArtiSynth, (b) Jaw model connected to a model of the tongue

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

- 3D finite element models, stably integrated with mass-spring and rigid body models and unilateral and bilateral constraints between rigid bodies (section 3);
- Collision detection between rigid bodies and finite element models (section 3);
- An enhanced airway including nasal passges and coupling to a 3D tongue model (section 3.3);
- Aeroacoustic improvements including: enhanced airway model with nasal passages, 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 is a platform to support the integration of different anatomical and acoustic models that have been created independently, such as those listed in section 2. We have created specific models with ArtiSynth such as 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 ariway/tongue model capable of synthesizing speech sounds (Doel et al., 2006). Figure 1 shows the jaw/laryngeal model, as well as the jaw model connected to the tongue, and the airway/tongue model is shown in Figure 3. Demos, information, and software can be obtained from the ArtiSynth web site: *www.artisynth.org*.

2. Related Work

Many researchers form 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.,

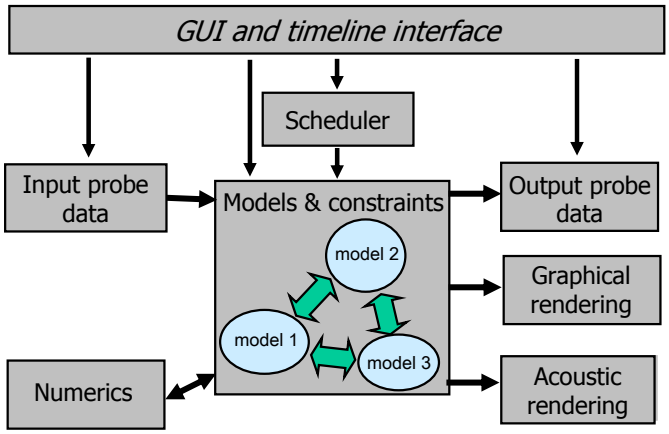

Figure 2. General architecture of the ArtiSynth system.

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).

These anatomical models are usually created by combining semi-automatic extraction methods and hand tuning procedures to generate model geometry and parameters (Gerard et al., 2003; McInerney and Terozopolis, 1999; Stavness et al., 2006) and use 1. engineering packages such as Femlab, Adams and ANSYS, 2. surgical simulation packages such as SOFA, GiPSi, SimTK and Open Tissue, and/or 3. movie and computer game physics effect tools and engines such as in Maya, Blender, RealFlow, openODE, and Havok for simulation.

The complex aero-acoustical processes that involve the interaction of these anatomical elements with airflow and pressure waves and which eventually produce speech have also been studied (Svancara et al., 2004; Sinder et al., June, 1998).

3. Modeling Capabilities

3.1. General Framework

The conceptual architecture for ArtiSynth is shown in Figure 2. At the center is the system model, which is a hierarchical arrangement of sub-models coupled together by *constraints* which mediate the interactions between them. Dynamic simulation of the system model is controlled by a *scheduler*, which calls on the sub-models to advance themselves through time, in the order required by their constraints, with the constraints adjusting model inputs and outputs. We have found that this model/constraint paradigm works for model connections which are parametric or non-stiff, but when a collection of models forms a stiff system, they must be combined into a single model (e.g., MechModel, Section 3.2) which advances the collection using a single implicit integrator.

Models and their components are implemented as Java classes, and provide support for naming, maintaining themselves within a hierarchy, reading and writing to persis-

tent storage, and (if desired) 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 that provide 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 can also be specified between rigid bodies, which permits the creation of joints and contacts. MechModel can be advanced dynamically using a choice of explicit or implicit integrators (forward Euler, Runga Kutta, backward Euler, or Newmark methods), although the presence of stiff components (such as FEMs) will usually require an implicit integrator.

3D FEMs are implemented using the linear stiffness-warping approach described in (Mueller and Gross, 2004), in which rotation is factored out of the strain tensor in order to reduce distortion. Our FEMs are currently tetrahedral, although an extension to hexahedral elements is expected. Muscle activation can be simulated within the FEM by applying uniform contraction forces along selected element edges, as described in (Vogt et al., 2006).

The primary components of MechModel which contain state are particles and rigid bodies, with the former serving as nodes in either mass spring models or FEMs. An individual particle or rigid body can be declared non-dynamic, in which case it's position and velocity is either fixed or controlled by model input or constraint. This allows parametric control of selected components within a MechModel.

Collision detection and handling between rigid bodies and FEMs is done by first using an oriented bounding box (OBB) method (Gottschalk et al., 1996) to find the interpenetrations between the surface meshes of the respective objects. Collisions between rigid bodies and FEMs are handled by projecting the penetrating FEM vertices onto the surface of the rigid body. Collisions between rigid bodies are handled by using the penetration contour to define a contact plane, and then applying a separating impulse in the direction normal to this plane. Collision handling between FEMs is still in development but is expected to entail mutual projection proportional to the FEM stiffness.

3.3. Acoustic Models

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 depicted in Figure 3. The mesh is structured as a number of cross-sections along the length of the tube, which

implicitly defines a center line. This structure permits fast calculation of the area function, which allows the acoustics to be modeled in a cylindrically symmetric tube. In addition to the surface mesh, the airway model allows for additional data such as labeling particular points or areas, or surface acoustical impedances, which may be needed for acoustical modeling within the tract. 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 which requires a fixed grid size (also available in ArtiSynth). The vocal chords are modeled using the

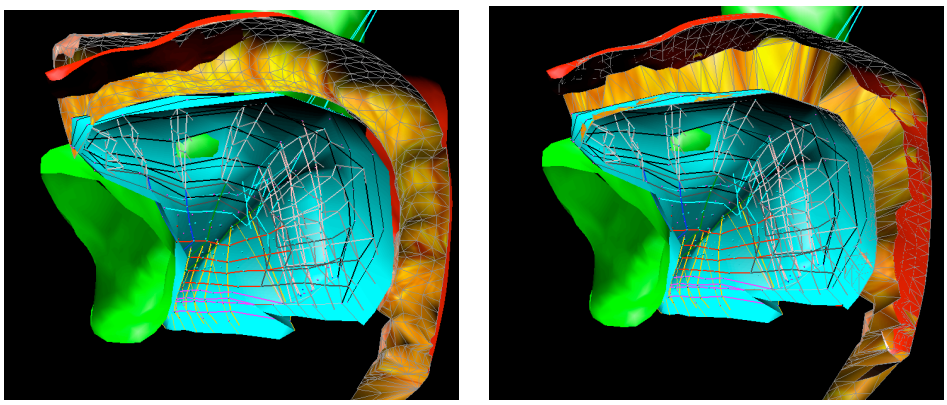

Figure 3. The Airway model and the tongue, palate and jaw meshes before (left) and after registration.

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 has Rosenberg excitation available 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 using 2 approaches: 1) directly in Java using appropriate method calls and 2) constructing an appropriate description in the ArtiSynth standard text-based file format for model components. The file format is also useful for the storing models which have been interactively edited, as described below.

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 model property values, transform geometry (translation, rotation, and scaling), and component addition and deletion.

5. Interactive Simulation

ArtiSynth provides means to "instrument" the 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.

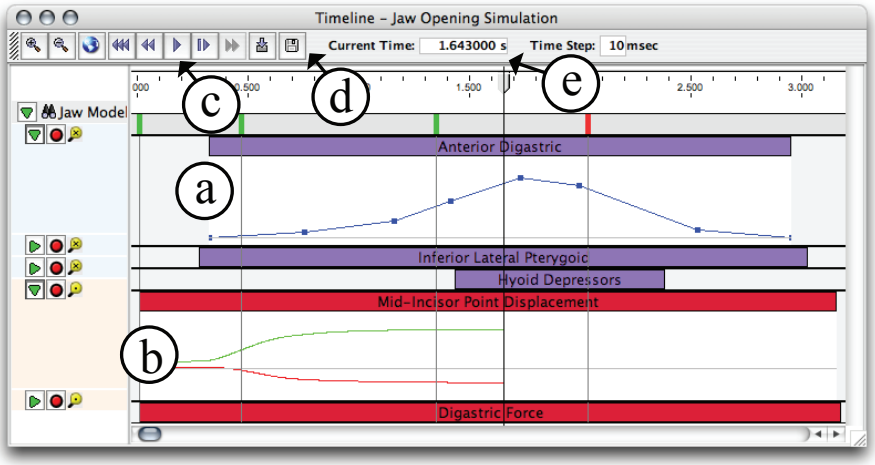

Figure 4. The ArtiSynth Timeline controlling a jaw motion simulation: (a) input probe (e.g. muscle activation), (b) output probe (e.g. incisor position), (c) play controls, (d) save / load buttons, (e) current time cursor.

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, showing control inputs arranged on the Timeline. Intuitively, our approach is to blend the timeline features of movie edit applications with the simulation environment concept from 3D modeling tools. ArtiSynth supports the use of *waypoints* to allow simulation states to be saved as the simulation progresses to allow researchers to go backward and forward in simulated time for dynamic models.

Attaching probes to a model can be done interactively, using a *probe edit* window. The edit window allows a user to specify exposed properties to which input or output probes should be attached. These properties can be selected using the navigation and selection mechanism mentioned in Section 4. Selected properties can be connected to numeric input or output channels associated with the probe, either directly or through simple numeric expressions.

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

The new features we have added to ArtiSynth derive from using it to create new, complex models of the jaw, tongue and combination of the two for speech synthesis. ArtiSynth's user-centered design strategy leads to a versatile, flexible and functional approach to physical modeling.

Acknowledgments

This work was supported by NSERC, Peter Wall Institute for Advanced Studies and the Advanced Telecommunications Research Laboratory (Japan). We gratefully acknowledge the many contributions made from the team of people contributing to this project 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.

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 Synthsizer. *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.