# An Efficient Biomechanical Tongue model for Speech Research.

Florian Vogt, John Lloyd, Stéphanie Buchaillard, Pascal Perrier, Matthieu Chabanas, Yohan Payan, Sidney Fels

# An Efficient Biomechanical Tongue model for Speech Research

**Florian Vogt**[1] **John E. Lloyd**[1] **Stéphanie Buchaillard**[2] **Pascal Perrier**[2]
**Matthieu Chabanas**[2] **Yohan Payan**[3] **Sidney S. Fels**[1]

[1] Dept. of ECE, University of British Columbia, Vancouver, Canada

[2]Institut de la Communication Parlée - INPG, 38031 Grenoble, France

[3]Laboratoire TIMC-GMCAO, Faculté de Médecine, 38700 La Tronche, France

***Abstract.*** *We describe our investigation of a fast 3D finite element method (FEM) for biomedical simulation of a muscle-activated human tongue. Our method uses a linear stiffness-warping scheme to achieve simulation speeds which are within a factor 10 of real-time rates at the expense of a small loss in accuracy. Muscle activations are produced by an arrangement of forces acting along selected edges of the FEM geometry. The model's dynamics are integrated using an implicit Euler formulation, which can be solved using either the conjugate gradient method or a direct sparse solver. To assess the utility of this model, we compare its accuracy against slower, but less approximate, simulations of a reference tongue model prepared using the FEM simulation package ANSYS.*

## 1. Introduction and Contributions

Basic speech motor control issues such as: the control of articulatory and/or acoustic trajectories shapes (Loefqvist and Gracco, 2002; Perrier et al., 2003; Zandipour et al., 2004), the control of intra- and inter -articulatory timing (Nam et al., 2006), or the control of prosodic factors (Fujimura, 2000) can be efficiently investigated by implementing and testing related hypotheses and control models on physical models of speech production. Then, simulations can be assessed via quantitative comparisons with experimental data allowing evaluation of models' adequacy. For such an approach to be efficient and

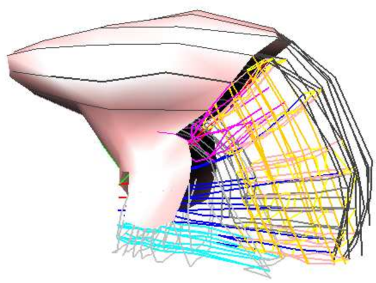


**Figure 1.** Tongue model, showing its surface mesh and (in cutaway) the FEM edges corresponding to muscle fibres

significant, there are two basic requirements. It is necessary (1) to assess a sufficiently large amount of simulations in order to take into account the whole variety of speech movements and speech sequences, and (2) to ensure that models are accurate enough to provide a fair representation of the physical speech production system. Intrinsically these two requirements are contradictory, in such a way that systematic compromises have to be found by reducing the number of simulations and/or by decreasing the realism of the modeling. In this context, elaborating new and original algorithms that could allow dramatic reductions of computation times without significant reduction of the modeling accuracy, is a major challenge. This is the aim of a collaborative project carried out at UBC Vancouver and ICP Grenoble, which aims at developing fast and accurate biomechanical models of speech articulators. This paper presents the first results that were obtained for a 3D biomechanical tongue model (Figure 1), originally developed in the standard Finite Element Package ANSYS (Gerard et al., 2006). This package provides accurate solutions, but at very large computational cost. We show that it is possible to rapidly calculate the dynamics of this model with reasonable accuracy, using a stiffness-warping technique such as that described in (Mueller and Gross, 2004). We have tested the accuracy of our faster approach by comparing it to the model computed using ANSYS. The contributions of our work include:

- Combining stiffness-warping with muscle forces acting along FEM edges to create a fast model of muscle-activated tissue;
- Demonstrating that this type of model can be integrated using an implicit integrator that can be solved with a conjugate gradient solver;
- Testing the accuracy of our approach against a reported reference tongue model

We anticipate that the techniques described here can be applied also to face and lip models.

This work is part of a larger project ArtiSynth, which is an open-source, Java-based, biomechanical simulation platform focused on the vocal tract and upper airway (Fels et al., 2006). One of the aims of ArtiSynth is to encourage collaboration and incremental development among scientific and medical researchers by making source code and model data easily available to the community.

## 2. Related Work

The tongue has been modeled in a wide variety of ways by various researchers. Parametric models of the tongue's shape have been developed using statistical methods (Badin et al., 1998; Engwall, 2000) and spline descriptions (Stone and Lundberg, 1996; King and Parent, 2001). An accurate physiological description is provided in (Takemoto, 2001). Dynamic models have been constructed using both discrete modeling approaches (Dang and Honda, 2004) and continuous finite element methods (Wilhelms-Tricarico, 1995; Payan and Perrier, 1997; Gerard et al., 2004, 2006). A recent survey (Hiiemae and Palmer, 2003) describes existing methods in detail.

In addition to being efficient to simulate (ideally at interactive rates) and being validated against real measurements, an effective tongue model must provide:

- Emulation of both tissue and muscle fibre;

- Large deformations, particularly at the tip;
- Incompressible and non-linear (hyperelastic) tissue deformation

Finite element methods provide a good solution for emulating both tissue and muscle fibre, and have a long tradition in Engineering (Zienkiewicz and Taylor, 2000). FEM models also provide greater stability and accuracy than mass-spring models; however, current FEM solutions (Wilhelms-Tricarico, 1995; Payan and Perrier, 1997; Gerard et al., 2004) do not compute in real-time. Recent developments in the fields of physical-based animation (Mueller and Gross, 2004; Teran et al., 2005) and surgical simulation (Cotin et al., 1999) provide finite element algorithms which can run in real-time, albeit with less (or even unknown) accuracy, to provide plausible results even for large deformations.

Most current approaches to muscle tissue modeling apply Hill's non-linear spring model (Hill, 1938) to either mass-spring systems (Dang and Honda, 2004), finite elements (Gladilin et al., 2001; Teran et al., 2005), or Cosserat models (Pai et al., 2005). For consistency with the reference tongue model we use a muscle model based on Feldman (1986) and Laboissiere et al. (1996).

## 3. Adaptation of the Reference Tongue Model Geometry

For this work, we use the 3D FEM tongue geometry developed in (Gerard et al., 2006), which is shown in Figure 1 and was developed from medical image data. This geometry contains 946 nodes, connected to form 740 hexahedral elements. These hexahedra were further subdivided into 3700 tetrahedra (using the optimal number of five tetrahedra per hexahedron) as our present implementation of the stiffness-warping algorithm requires tetrahedral geometry.

We created a fast deformable model of the FEM tongue by adapting the stiffness-warping FEM approach developed by Mueller and Gross (2004) to handle muscle activations. Details about our implementation can be found in Vogt et al. (2006).

## 4. Implementation

Our tongue model is implemented in Java, using the modeling and numeric library support of ArtiSynth (Fels et al., 2006), which provides a framework for creating and interconnecting various kinds of dynamic and parametric models to form a complete integrated biomechanical system.

Since execution speed is an important issue, it should be mentioned that the dynamic native code compilers (e.g., Hotspot) provided by current Java implementations usually produce execution speeds that approach compiled C/C++ code (Nikishkov, 2003).

To solve the system, we used either the Pardiso sparse solver (Schenk et al., 2003), or a conjugate gradient (CG) method; further details are given in Section 6.

## 5. Results: Accuracy

In this section we compare the accuracy of our stiffness-warping FEM implementation (denoted as WRP) with two methods implemented using the industry standard FEM pack-

| Task | Muscle activations | LSD deviation (mm) | | WRP deviation (mm) | |
|---|---|---|---|---|---|
| | | max | mean | max | mean |
| A | posterior genioglossus (2.0N) | 2.3 | 1.0 | 1.3 | 1.0 |
| B | anterior genioglossus (0.5N) | 3.2 | 1.6 | 1.0 | 0.9 |
| C | hyoglosse (2.0N) | 3.3 | 1.8 | 1.1 | 0.9 |
| D | transversalis (2.0N) | 1.2 | 0.6 | 0.6 | 0.4 |
| E | inferior longitudinalis (0.5N) | 1.4 | 1.1 | 0.9 | 0.8 |

**Table 1.** Muscle activation tasks and end-task deformation errors of these tasks (compared to HYP), resulting from the methods LSD and WRP.

age ANSYS: (1) a linear small-deformation model (LSD), and (2) a hyperelastic Mooney-Rivlin solid model (HYP). Note that we implemented the LSD model in ArtiSynth and got identical results. All models used the same tetrahedral meshing described in Section 3.

The tissue elasticity parameters were obtained from the experimental work reported by Gerard et al. (2004). For the WRP and LSD models, we used a Young's modulus of $E = 6912$ and a Poisson's ratio of $\nu = 0.49$. For HYP, we set $C_1 = 1152$, $C_2 = 540$, and $\nu = 0.49$. All models used a Raleigh damping (i.e., $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$) with $\alpha = 6.22$ and $\beta = 0.11$.

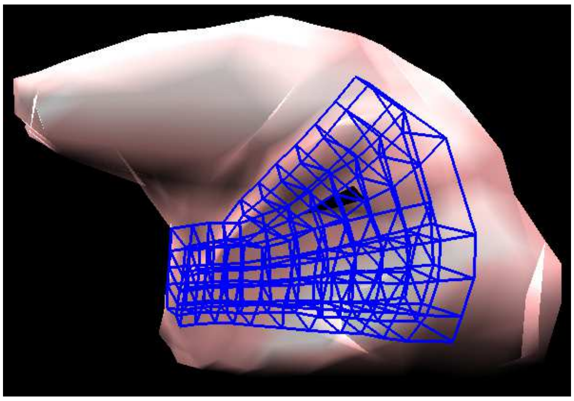
These models were used to simulate a set of five tasks in which a constant excitation was applied to one or more tongue muscles for 1.2 seconds and observed at a rate of 10 ms. The tasks are named and described in Table 1. The HYP and LSD models were computed using a variable rate ANSYS integrator, while our WRP model was computed using the single step implicit integration scheme of Vogt et al. (2006) with a fixed time step of 10ms.

To assess model accuracy, the deformations resulting from WRP and LSD were compared against those of the HYP (which was considered to be the most accurate and so was used as a reference). Specifically, the deformations $\mathbf{u}_i$ of a set of ten nodes lying on the tongue's mid-sagittal plane were compared against the reference deformations $\mathbf{u}_{ri}$ resulting from the HYP model. The deformation error $e_i$ at each sample point was then computed simply as

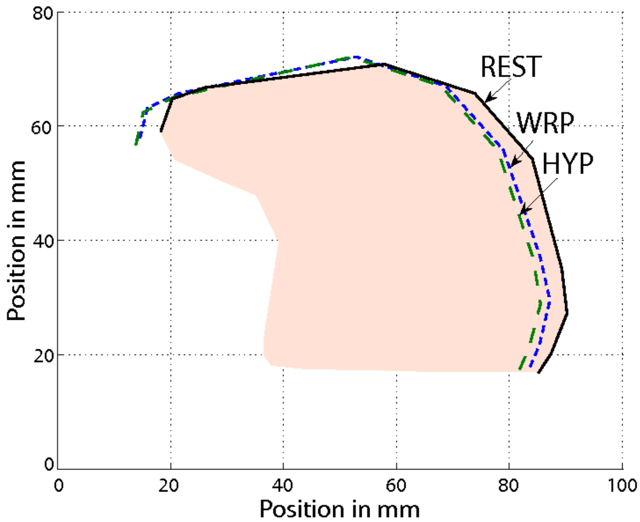
$$e_i = \|\mathbf{u}_{ri} - \mathbf{u}_i\|. \tag{1}$$

The mean and maximum of $e_i$ were used to gauge the overall deformation error. Table 1 shows these values for both the WRP and LSD at the end points of each of the tasks. Figure 2(a) shows the unactivated tongue model, while 2(b) shows the mid-sagittal plane nodes before activation, and after activation for Task A as modeled by both HYP and WRP.

Figure 2(b) suggests that the deformations produced by our stiffness-warping model do in fact adhere quite closely to those produced by the hyper-elastic reference model. This is supported more quantitatively by Table 1, where the mean error for WRP is alway within 1 mm, and the maximum error is close to this as well, which is 2 to 3 times better than the results for LSD. Ideally, we could consider the stiffness-warping model to



(a)




(b)

**Figure 2.** (a) Tongue at rest before activation, showing the fibres of the posterior genioglossus muscle; (b) Nodes in the tongue's mid-sagittal plane before activation (REST) and after activation of Task A, as modeled by both WRP and HYP

match perfectly the WRP model, if the maximum error would be smaller than 0.5mm. Indeed, experimental works on articulatory variability (see for example (Ma et al., 2006) in this conference) suggest that differences of the order of 1mm in tongue positionning could be the correlates of coarticulatory strategies. Further improvements of the model should allow us to get close to this ideal objective.

## 6. Results: Speed and Stability

Computation times for the results reported above were markedly faster for the WRP model as compared with the ANSYS LSD and HYP models, with the former requiring only about 10 CPU seconds per simulated second, while ANSYS required about 600 CPU seconds. All tests were run on a 2.8GHz Pentium IV single processor computer.

For the implicit integration step, we used the Pardiso spare solver (Schenk et al., 2003). We also used a conjugate gradient (CG) solver, although this was slower than Pardiso because we did not use a preconditioner and so about 300 iterations were required to achieve equivalent accuracy.

With regard to stability, we found that our implicit solver was stable at time steps of up to 20ms, assuming our implicit scheme muscle activations (Vogt et al. (2006)). Larger time steps lead to instabilities which cause severe loss of accurcy. Using explicit methods for muscle handling causes the stability limit to drops to around 5ms.

## 7. Conclusion and Future Work

In this work we introduce a fast and stable finite element model and compare its performance to ANSYS FEM simulation using a previously published reference tongue model. The comparison required the adaptation of the hexahedral tessellation of the reference tongue model to a tetrahedral tessellation. Our method admits simulations speeds that are within a factor 10 of real-time, at the expense of a small loss in model accuracy.

Further evaluation of out fast finite element tongue model consists in applying complex muscle activation patterns similar to those underlying the production of speech sequences. Vowel-vowel and vowel-consonant-vowel sequences will thus be generated and the corresponding movements will be compared to those produced with the ANSYS model, not only in terms of tongue positioning accuracy at the targets, but also in terms of spatial trajectory shapes and tangential velocity profiles.

These results will be presented at the conference.

## References

Badin, P., Bailly, G., Raybaudi, M., and Segebarth, C. A three-dimensional linear articulatory model based on mri data. In *Proceedings of the International Conference of Spoken Language (ICSLP)*, pages 14–20, 1998.

Cotin, S., Delingette, H., and Ayache, A. Real-time elastic deformations of soft tissues for surgery simulation. *IEEE Trans. Vis. & CG.*, 5(1):62–73, January-March 1999.

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

Engwall, O. A 3D tongue model based on MRI data. In *Proceedings of the International Conference of Spoken Language (ICSLP)*, 2000.

Feldman, A. Once more on the equilibrium-point hypothesis model for motor control. *Journal of Motor Behavior*, 18(1):17–54, 1986.

Fels, S., Vogt, F., van den Doel, K., Lloyd, J., Stavness, I., and Vatikiotis-Bateson, E. Artisynth: A biomechanical simulation platform for the vocal tract and upper airway. Technical Report TR-2006-10, Computer Science Dept., University of British Columbia, 2006. URL `http://www.artisynth.org`.

Fujimura, O. The C/D model and prosodic control of articulatory behavior. *Phonetica*, 57(2-4):128–138, 2000.

Gerard, J., Ohayon, J., Luboz, V., Perrier, P., and Payan, Y. Indentation for estimating the human tongue soft tissues constitutive law: application to a 3d biomechanical model to study speech motor control and pathologies of the upper airways. *LNCS*, 3078:77–83, 2004.

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

Gladilin, E., Zachow, S., Deuflhard, P., and Hege., H.-C. Virtual fibers: A robust approach for muscle simulation. In *Proc MEDICON*, pages 961–964, 2001.

Hiiemae, K. M. and Palmer, J. B. Tongue movements in feeding and speech. *Crit Rev Oral Biol Med*, 14:430–449, 2003.

Hill, A. The heat of shortening and the dynamic constants of muscle. *Proc Roy Soc B*, 126:136–195, 1938.

King, S. A. and Parent, R. E. A 3d parametric tongue model for animated speech. *JVCA*, 12(3):107–115, 2001.

Laboissiere, R., Ostry, D. J., and Feldman, A. G. The control of multi-muscle systems: human jaw and hyoid movements. *Biological Cybernetics*, 74:373–384, 1996.

Loefqvist, A. and Gracco, V. Control of oral closure in lingual stop consonant production. *JASA*, 111(6):2811–2827, 2002.

Ma, L., Perrier, P., and Dang, J. Anticipatory coarticulation in vowel-consonant-vowel sequences: A crosslinguistic study of french and mandarin speakers. In *Submitted to ISSP*, 2006.

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

Nam, H., Goldstein, L., and Saltzman, E. Dynamical modeling of supragesturaltiming. In *Proc. of LabPhon10, Paris*, 2006.

Nikishkov, G. Java performance in finite element computations. In *Proc Appl Sim & Mod*, page 410, 2003.

Pai, D. K., Sueda, S., and Wei., Q. Fast physically based musculoskeletal simulation. *ACM Trans Graph*, 2005.

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

Perrier, P., Payan, Y., Zandipour, M., and Perkell, J. Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *JASA*, 114(3):77–83, 2003.

Schenk, O., Röllin, S., and Hagemann, M. Recent advances in sparse linear solver technology for semiconductor device simulation matrices. In *IEEE SISPAD*, pages 103–108, 2003.

Stone, M. and Lundberg, A. Three-dimensional tongue surfaces from ultrasound images. In *SPIE Proc.*, pages 168–179, 1996.

Takemoto, H. Morphological analysis of the human tongue muscularture for three-dimensional modeling. *J. Sp. Lang. Hear. Res.*, 44:95–107, 2001.

Teran, J., Sifakis, E., Blemker, S., Ng Thow Hing, V., Lau, C., and Fedkiw, R. Creating and simulating skeletal muscle from the visible human data set. In *IEEE TVCG (in press).*, 2005.

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.

Wilhelms-Tricarico, R. Physiological modeling of speech production: methods for modeling soft-tissue articulators. *JASA*, 97(5):3085–98, 1995.

Zandipour, M., Guenther, F. H., Perkell, J. S., Perrier, P., Payan, Y., and Badin, P. Vowel-vowel planning in acoustic and muscle space. In *Proc. of In From Sound to Sense: 50+ years of discoveries in speech communication*, pages C103 – C108, 2004.

Zienkiewicz, O. and Taylor, R. *The finite element method*. Oxford, 2000.