Human Perioral Dynamics Model for Labial Interaction in Speech Articulation and its Motor Control

Kangsoo Kim[1], Hiroaki Gomi[1†]

[1]NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation
3-1, Wakamiya, Morinosato, Atsugi, Kanagawa, 243-0198, Japan

kskim@idea.brl.ntt.co.jp, gomi@idea.brl.ntt.co.jp

Abstract. *We investigated a computational model of labial interaction dynamics and motor control in human speech articulation. Simulation is carried out using a dynamic model of human articulators consisting of lip soft tissues with related muscles and bones. The continuum of lip soft tissue is modeled as a discrete approximation composed of networked point masses interlinked by viscoelastic elements. Stiffness of viscoelatic elements is adjusted to ensure the compatibility in static deformation between discrete model and its continuum prototype. Incompressibility of human facial tissue is incorporated into the dynamic model by introducing an artificial actuation. To prevent from the mutual penetration of upper and lower lips, collision response scheme based on penalty actuation has been developed. As a mathematical description of the human speech acquisition, iterative method for estimating neuromuscular commands has been introduced. Human lip movements during speech articulation were successfully mimicked by the estimated motor command.*

1. Introduction

This paper addresses a simulation based investigation of human articulatory motion, in view of the dynamic deduction and acquisition of the speech control strategy.
In our simulation model, discrete model representation is used approximating real human skin and soft tissue continuum. This discrete model representation has been employed in several former researches (Terzopoulos et al., 1993; Dang et al., 2001). If the elements in a discrete model are distributed with uniform volumetric density, constant element stiffness makes the discrete model compatible with continuum in static deformation. But except for the extremely simple geometrical structure, it is hardly possible to construct a discrete model of uniform element density. In this research, a new technique of adjusting element stiffness is proposed, which lets the discrete model of arbitrary element density exhibit continuum compatible static deformation.
Real human skin and soft tissue is known to be incompressible. Lee and Terzopoulos (1995) proposed volume preservation force made of static restoring of volume change and nodal displacement. Dang and Honda (2001) suggested the description of constant volume constraint on the basis of the semicontinuum assumption. The concept of the cylinder, which is the artificial volumetric body of viscoelastic elements, is introduced to make their method to work. In this research, a volume regulating actuation acting on

[†]Supported by Shimojo Implicit Brain Function Project, Japan Science and Technology Agency.

the outer boundary surface of the lip soft tissue model is presented to realize the constraint of incompressibility. The actuation is composed of the feedback of current volume error and volumetric flux.
During the speech articulation, it sometimes happens that upper lip collides with lower one, resulting in the deformed lip shape to reach an equilibrium position. We present a simple technique of collision detection and treating lip deformation due to collision.
It is generally accepted that during the language acquisition, a human being learns about auditory-articulatory relationships under feedback control (Harrington and Tabian (2005)). In this research, inverse dynamics problem in speech articulation is treated as a mathematical description of the human speech acquisition. Similar to Pitermann and Munhall (2001), we estimated the neuromuscular commands for speech articulation using an inverse estimation method. While they implemented eight pairs of facial muscles controlled symmetrically, our model permits unsymmetrical activation of the paired ones in total twelve kinds of muscles (Figure 1). In comparison to the frame based iteration by Pitermann et al. (2001), our estimation is updated with forward dynamics proceeding in parallel, leading to the inclusion of dynamic hysteresis in inverse estimation.

2. Human Articulator Dynamic Model

2.1 Dynamic Model Description

Figure 1 shows the configuration of the articulator dynamic model used as the plant for articulatory dynamics and control simulation. Articulators represented in the dynamic model are lip soft tissue with related muscles and bones. The articulator dynamic model presented in this work has thirty-three sets of muscles and two sets of bones. Muscle Structure is the one presented by Gomi and Nozoe et al. (2005), constructed on the basis of anatomic reality.

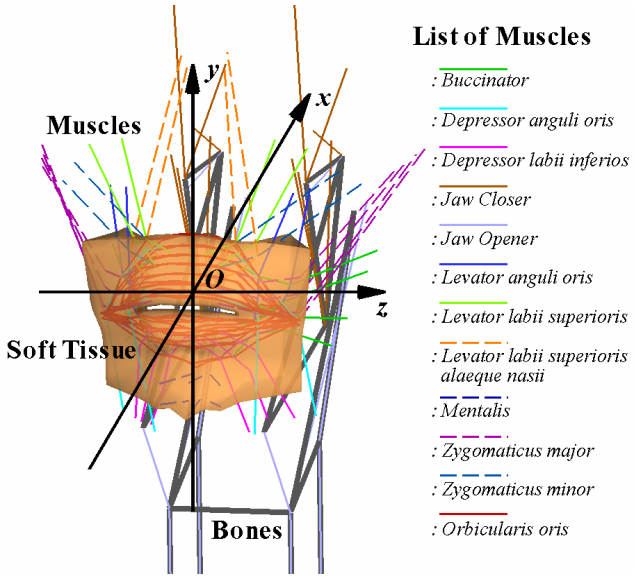


Figure 1. Configuration and coordinate system of the human articulator dynamic model.

2.2 Continuum Compatible Discrete Model Approximation

Human skin including inside soft tissue is made of a continuum. In this research, we derive a discrete model of human lip soft tissue and employ it as a component of the articulator dynamic model. Discrete lip soft tissue model consists of a set of lumped point masses interlinked by viscoelastic elements. The viscoelastic element is a mechanical abstraction consisting of a spring and a dashpot combined in parallel. In Figure 2(b,c), each line segment represents a viscoelastic element unit.

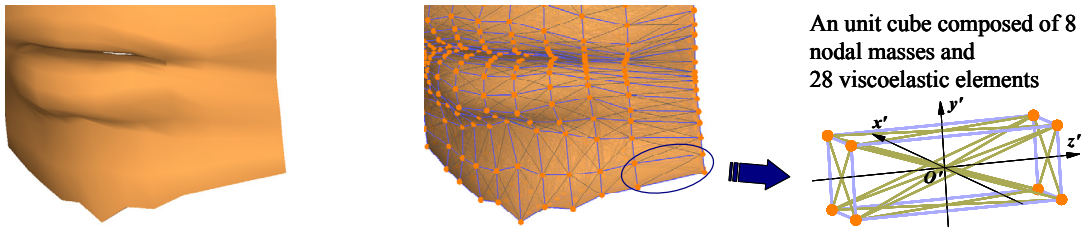


Figure 2. Computational model representation of lip soft tissue: (a) Lip soft tissue continuum. (b) Discrete model approximation of lip soft tissue continuum. (c) An unit cube in discrete model.

In order to derive a continuum compatible discrete model of lip soft tissue, we propose a method that optimally adjusts the stiffness of viscoelastic elements. For our adjustment to work, entire structure of the model is considered to be an assemblage of "unit cubes" shown in Figure 2c. Since we can easily calculate the static deformation of a continuum, stiffness adjustment is carried out to satisfy the followings: for the equivalent load condition, forced deformation of unit cube approximates that of geometrically identical continuum with minimal deviation. This adjustment is applied with respect to local x,y and z directions defined within each unit cube.

2.3 Equations of Motion

Equations of motion are derived by linear momentum conservation on each point mass. Since the point masses are interconnected with each other via viscoelastic elements, a coupled set of equations of motion for multi-DOF system is obtained as :

$$\boldsymbol{m}\ddot{\boldsymbol{R}} = -\boldsymbol{K}(\boldsymbol{R}-\boldsymbol{R}_0) - \boldsymbol{C}\dot{\boldsymbol{R}} + \boldsymbol{F}_a \qquad (1)$$

where $\boldsymbol{m}$ is the nodal mass matrix and $\boldsymbol{R}$ ($\boldsymbol{R}_0$) is the current (initial) nodal position vector. Matrices $\boldsymbol{K}$ and $\boldsymbol{C}$ represent stiffness and damping of the model. Vector $\boldsymbol{F}_a$ is the summation of external actuations, such as muscular force or volume regulating force. The damping matrix $\boldsymbol{C}$ is obtained by the exactly same way as the $\boldsymbol{K}$, except that the element stiffness is substituted for the damping of corresponding element.

2.4 Incompressibility of Lip Soft Tissue

In order to properly include the incompressibility condition within the human soft tissue dynamics, few methods have been reported. Dang and Honda (2001) let the volume change of hexahedral meshes combine with the equations of motion, by assigning it for the Lagrangian. Lee and Terzopoulos (1995) introduce volume preservation restoring

force made of the change in meshed volume and nodal displacement. To satisfy the constant volume constraint of human lip soft tissue, we introduce an auxiliary actuation regulating the model generating volumetric flux toward zero. Similar to the fluid flux over a control surface, volumetric flux of the soft tissue model is calculated by the integration of surface normal velocity over the entire outer boundary surface of it :

$$Q = \int_{S_{out}} \boldsymbol{V}_S \cdot \boldsymbol{n}\, dS \tag{2}$$

S_{out} represents the outer surface of lip soft tissue model, $\boldsymbol{V}_S$ the nodal velocity on S_{out}, and $\boldsymbol{n}$ the unit normal vector on S_{out}. Positive $\boldsymbol{n}$ is defined to direct outward from the region of lip soft tissue. In addition to the volumetric flux, time integration of it, the current volume error, is also made to contribute the volume regulating actuation.

$$\boldsymbol{f}_{vol\pm} = \mp\left(K_p Q + K_I \int_0^t Q(\tau)\, d\tau \right) \boldsymbol{n} \tag{3}$$

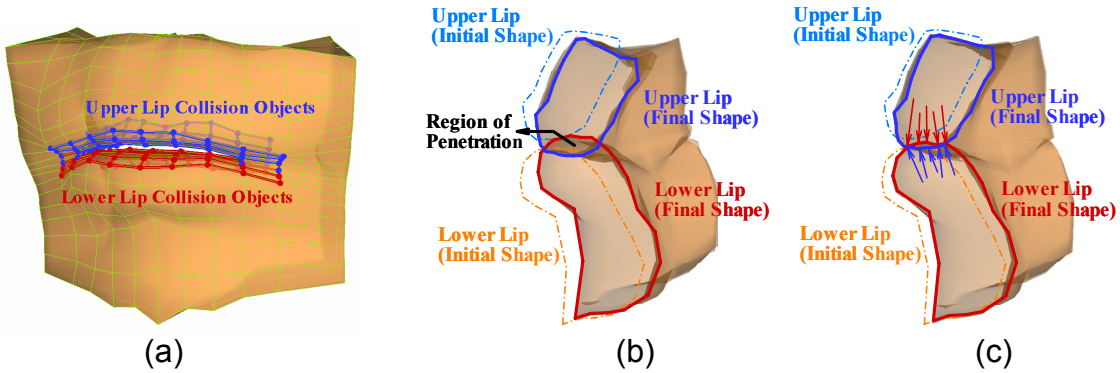
The actuation presented here is made of PI (proportional-plus-integral) compensation of volumetric flux, expected to suppress the volumetric flux more effectively compared to the volume error feedback alone. And since $\boldsymbol{f}_{vol\pm}$ is evaluated and applied only on the outer boundary surface, it requires relatively low cost of computation.

2.5 Lip Soft Tissue Collision

During speech articulation, it frequently happens that the collision between upper and lower lips. Typical examples are the articulatory motions to produce the speech acoustics based on the consonants such as /b/, /m/ or /p/. Similar to the case of incompressibility condition, lips collision response is modeled by introducing an extra actuation named collision penalty force. The main purpose of collision response modeling in our dynamic model is to prevent from the mutual penetration between upper and lower lips. The collision penalty force is applied on the lip nodes experiencing the collision with opposite side lip. In detecting the lips collision, instead of pursuing rigorous detection, we extract the colliding node set by just examining the distance between a lip surface node and every collision object included in the opposite side lip. The collision objects are collection of nodes and patches consisting of 4-nodes on its vertices (Figure 3(a)).
When the collision is detected, penalty force defined as follows is applied on the colliding node set to prevent from penetrating across the opposite side lip surface.

$$\boldsymbol{f}_{CPi} = -\frac{c_p}{l_{ij}^2} \boldsymbol{v}_i \tag{4}$$

For a colliding node i, l_{ij} represents the distance between the node i and the colliding object j in the opposite side lip. $\boldsymbol{v}_i$ is the nodal velocity and c_p the collision damping.
Figure 3(b,c) shows the simulated collision response of lip soft tissue projected on the midsagittal section. As shown in the figure, while penalty force leads to the virtual equilibrium in position and deformation after the collision occurrence, the collision response without penalty force suffers severe unrealistic penetration.



(a) (b) (c)

Figure 3. (a) Collision objects (Nodes and Patches) (b) Collision response without penalty force (c) Collision response with penalty force

3. Volume Conservation during Lip Motion Simulation

Figure 5 demonstrates the sequence of volume error and volumetric flux during the pseudo-speech motion of speaking /u/ (Figure 4). Results are taken from two conditions of the simulation, with (Figure 5(b)) or without (Figure 5(a)) the volumetric flux regulating actuation $\boldsymbol{f}_{vol\pm}$. As shown in the figure, $\boldsymbol{f}_{vol\pm}$ obviously reduces the volumetric flux at every time step, resulting in the volume error at final time *0.3*(%) of entire lip soft tissue volume. Without $\boldsymbol{f}_{vol\pm}$, final volume error reaches up to *9.5*(%).

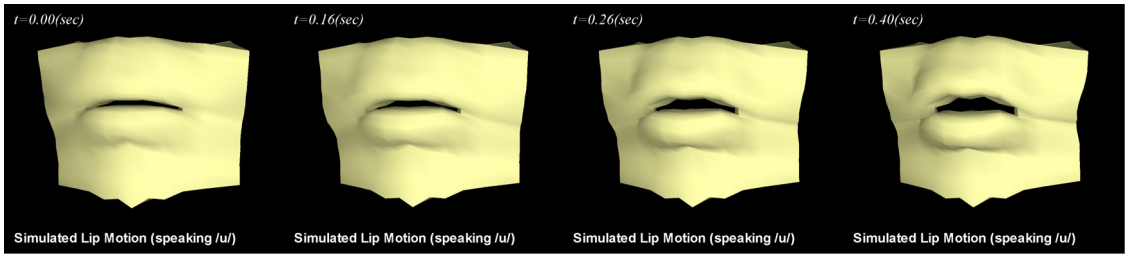
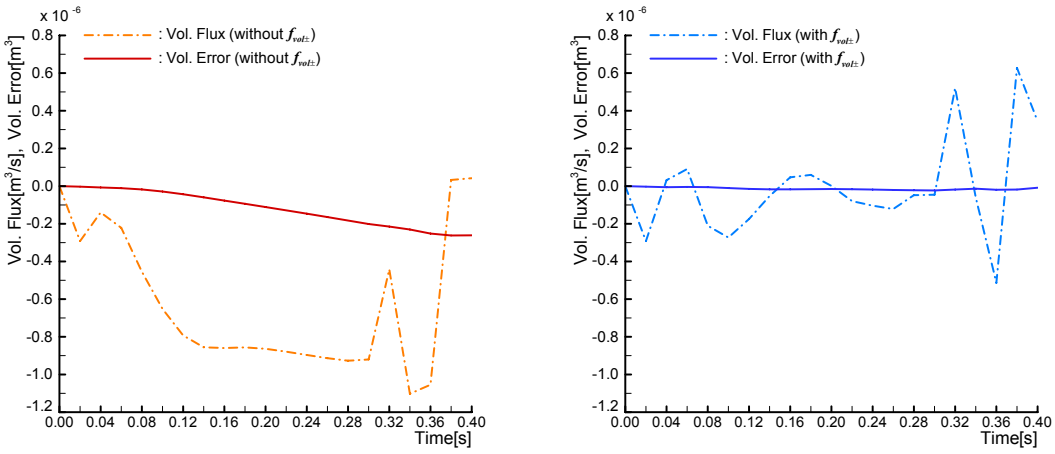


Figure 4. Pseudo-speech lip motion speaking /u/



(a) (b)

Figure 5. Volumetric flux and volume error during the simulations (a) without vol. flux regulating actuation (b) with vol. flux regulating actuation

4. Inverse Dynamics Based Articulatory Motion Control

4.1 Estimation of Neuromuscular Command in Human Speech Articulation

In an inverse dynamics problem, a desired output is given and the problem is to determine the input required to produce the output. Thus the problem of inverse dynamics in human speech articulation is the estimation of neuromuscular command to produce a target speech acoustics that the speaker intends to produce. But since our dynamic model does not actually generate speech acoustics, as the reference for the inverse dynamics problem, kinematic motion responses of the lip articulator corresponding to a reference speech acoustics is used instead.
Achievement of estimation is evaluated by the performance index J defined as follows.

$$J = \frac{1}{2}\left\| \boldsymbol{x}_{ref} - \boldsymbol{x} \right\|^2 \tag{5}$$

$\boldsymbol{x}_{ref}$ and $\boldsymbol{x}$ represents the reference and the estimated position vector of selected points within the model, called observation points.
The algorithm of motor command estimation used here is the Gradient Descent Search (GDS). Using GDS, Gomi et al. (2004) solved the inverse dynamics of tongue motion. According to their work, ,GDS updates the neuromuscular command as Equation (6) at each iteration step.

$$\boldsymbol{\alpha}_{i+1} = \boldsymbol{\alpha}_i + \Delta\boldsymbol{\alpha} \tag{6}$$

$\boldsymbol{\alpha}_i$ represents the neuromuscular command vector at iteration step i, while $\Delta\boldsymbol{\alpha}$ its correction. The application of GDS derives the correction vector $\Delta\boldsymbol{\alpha}$ as defined Equation 6, which makes the $dJ/d\tau$, derivative of performance index with respect to iteration, negative semidefinite.

$$\Delta\boldsymbol{\alpha} = \eta \frac{d\boldsymbol{\alpha}}{d\tau} = \eta\left[(\boldsymbol{x}_{ref} - \boldsymbol{x})\boldsymbol{G} \right]^T \tag{7a}$$

$$\boldsymbol{G} = \frac{d\boldsymbol{x}}{d\boldsymbol{V}} \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{F}_a} \frac{\partial \boldsymbol{F}_a}{\partial \boldsymbol{\alpha}} \tag{7b}$$

Therefore, iterative update of $\boldsymbol{\alpha}_i$ by Equation 6 reaches the converged solution with minimized performance index. In Equation 7, η represents the gain, $\boldsymbol{V}$ the nodal velocity and $\boldsymbol{F}_a$ the external actuation.

4.2 Results of Estimation

The reference lip motion is taken from the actual speech articulation performed by a Japanese male subject. 3-D position and velocity of the lip motion is measured by a visual motion capture system.
To measure the reference kinematics by the motion capture system, markers are attached on specified observation points (Figure 6).

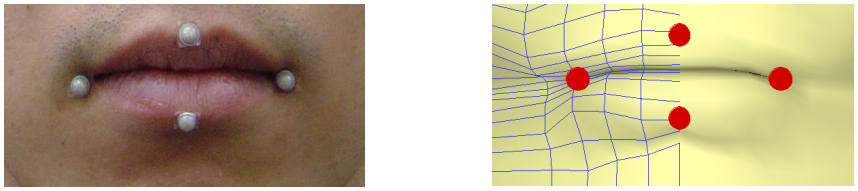


Figure 6. Markers on observation points (subject-fixed / simulation model)

As the reference, articulatory motion of speaking /aba/ performed by the subject is measured. In Figure 7, estimated motion is compared to the measured reference. Starting from the neutral position, estimation recovers the reference with minimized residual error after 8 times of iterations in GDS.

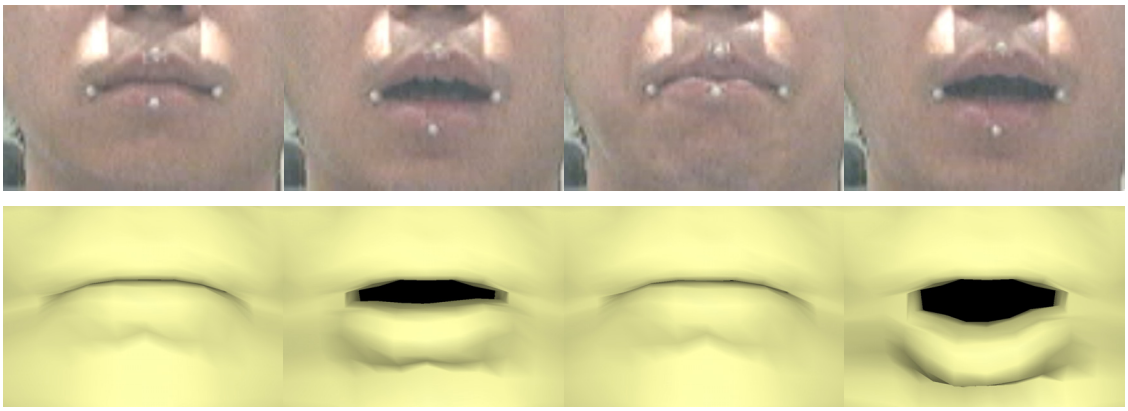
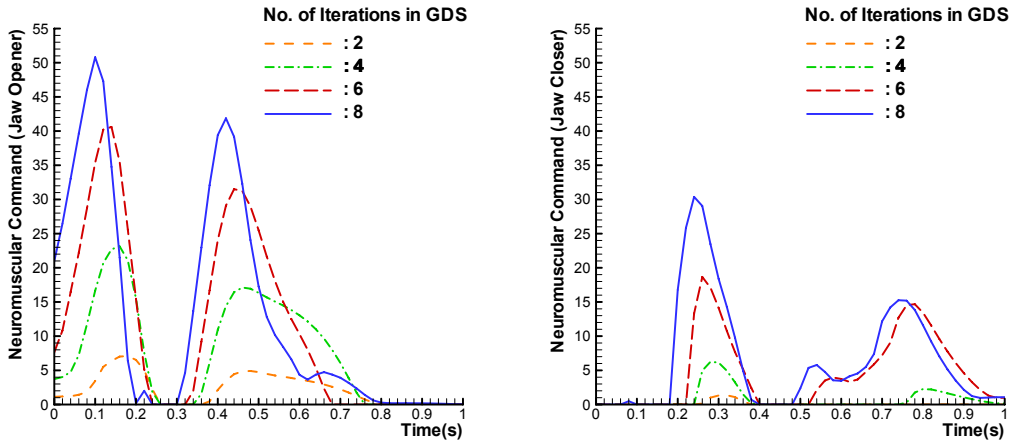


Figure 7. Lip motion sequence of speaking /aba/ : reference (top) vs. estimated (bottom) at *t=0.0*, *0.2*, *0.34* and *0.6* (sec), respectively

Figure 8 shows the estimated neuromuscular commands. Commands for two kinds of muscles show significant activation to induce the estimated lip motion. The jaw opener is activated during 0.0 ~ 0.24 (sec) and 0.28 ~ 0.8 (sec), each of which corresponds to the induction of lip opening behavior during the estimated motion of speaking /aba/. Activation of jaw closer is responsible for the lip closing motion followed by the lip opening one.



(a) (b)

Figure 8. Transition of estimated neuromuscular commands inducing lip motion sequence of speaking /aba/ (a) Jaw Opener (b) Jaw Closer

5. Conclusion and Future Works

By activating the necessary neuromuscular command obtained from inverse dynamics based estimation, we succeeded in making the human perioral dynamic model mimic the lip motion for actual speech articulation. The dynamic model is the one enhanced to realize the labial interactions in speech articulation, such as lip soft tissue collision and volume conserving motion response. The discrete model of lip soft tissue is the one possessing continuum compatibility in static deformation. In order to reach the proper estimation of dynamic model generating lip motions in view of speech articulation, availability of the biophysically reasonable constraints seems one of the key issues. Finally, though we have treated the dynamic model of lips only, integration with the model of tongue and vocal tract will present advanced tool for investigation of human speech articulation.

References

Dang, J. and Honda, M. A physiological model of a dynamic vocal tract for speech production. *Acoust. Sci. and Tech.*, (22)6:415-425, 2001.

Gomi., H., Nozoe, J., Dang, J., and Honda, K. Physiologically based lip model for generating speech articulation. In *Proceedings of 6th International Seminar on Speech Production*, Manly, Sydney, Macquarie center for cognitive science, 2003.

Gomi., H., Nozoe, J., Dang, J., and Honda, K. A physiologically based model of perioral dynamics for various lip deformations in speech articulation. In Harrington, J. and Tabain, M., editors, *Speech Production: Models, Phonetic Processes, and Techniques*, pages 119-134. Psychology Press, 2006.

Gomi, H., Piao, X., Nozoe, J., Dang, J., and Honda, M. Movement control of articulatory dynamics model by imitation learning. In *Technical Report of IEICE,* NC2003-124, pages 31-36, 2004 (in Japanese).

Harrington, J. and Tabain, M. *Speech Production: Models, Phonetic Processes, and Techniques* In Harrington, J. and Tabain, M., editors, Psychology Press, 2006.

Lee, Y., Terzopoulos, D., and Waters, K. Realistic modeling for facial animation. In *Proceedings of SIGGRAPH95, ACM SIGGRAPH*, pages 55-62, Los Angeles, CA, U.S.A., Aug.1995.

Lucero, J.C. and Munhall K.G. A model of facial biomechanics for speech production. *J. Acoust. Soc. Am.*, 106(5):2834-2842, 1999.

Pitermann, M. and Munhall K.G. An inverse dynamics approach to face animation. *J. Acoust. Soc. Am.*, 110(3):1570-1580, 2001.

Terzopoulos, D. and Waters, K. Analysis and synthesis of facial image sequences using physical and anatomical models, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 15(6):569-579, 1993.