TR-H-247

A Physiological Model of a Dynamic Vocal Tract for Speech Production.

Jianwu DANG and Kiyoshi HONDA

1998.6.22

ATR人間情報通信研究所
〒619-0288　京都府相楽郡精華町光台2-2　TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
Telephone: +81-774-95-1011
Fax : +81-774-95-1008

A Physiological Model of a Dynamic Vocal Tract for Speech Production

Jianwu DANG and Kiyoshi HONDA

ATR Human Information Processing Research Labs,

2-2 Hikaridai Seika-cho Soraku-gun, Kyoto, Japan, 619-02

Abstract

A physiological articulatory model has been developed to simulate the dynamic actions of the speech organs. This model consists of the tongue, mandible, hyoid bone, and the outer wall of the vocal tract in three dimensions. The soft tissue of the tongue is outlined by a midsagittal plane and two parasagittal planes with 1-cm intervals. The mandible and hyoid bone are modeled to yield rotation and translation motion. The palatal and pharyngeal walls are modeled as a rigid structure. Thus, the proposed dynamic model of the vocal tract consists of soft and hard components with the rigid outer wall of the vocal tract. The frameworks of the rigid organs and muscle geometry of the model are extracted from volumetric MR images of a male speaker. All the soft tissue and rigid organs are modeled using mass-points and two types of connections: viscoelastic springs for connective tissue, and rigid beams for bony organs. The muscle definitions are based on an improved Hill's model. This study also develops a model's control method to generate tongue and jaw movements of the model by using generalized muscle activation signals that are similar to integrated EMG signals. The collisions of the tongue and the outer wall of the vocal tract were also considered in the model. The model produces plausible dynamic patterns of the tongue-jaw complex with relatively short computational time in comparison with the finite element method because the displacements of both the soft tissue and rigid organs are computed with the same algorithm.

PACS number: 43.70.Aj, 43.70.Bk

INTRODUCTION

The understanding of the detailed mechanisms of speech production is important for the studies of speech communication because anatomical and biomechanical properties of the speech organs are believed to have significant influences on articulatory kinematics, speech motor control strategy, sound patterns of languages, and phonological processes. However, it is impossible to experimentally reveal all details of the human speech organs, and thus a model of human speech organs helps us understand certain characteristics of human speech production. To this end, various models have been developed to simulate the mechanisms of speech production and examine the influences of the anatomical and biomechanical properties on speech. A number of geometrical or parametrical articulatory models have been developed to describe the relationship of articulation and acoustic signals (cf. Hence, 1967; Mermelstein, 1973; Coker, 1976; Maeda, 1990; Sondhi and Schroeter, 1987; Saltzman and Munhall 1989). In these models, the anatomy and physiology of the tongue are not adopted in computer simulation, although some attempts have been reported. On the other hand, physiologically-based articulatory models aim to recapitulate physiological functions of the speech organs. Several researchers have developed this type of models for the major articulators individually, such as the tongue, jaw, and lips.

Perkell's work (1974) was the first to construct a physiological model of the human tongue. His tongue model is a two-dimensional projection of the tongue in the sagittal plane and composed by a lumped parameter and lumped force system, which has some resemblance to the finite element method (FEM). The finite element method has been applied to the 3D tongue models by Kiritani et al. (1976), Kakita et al. (1985), and Hashimoto and Suga (1986). These models are essentially based on infinitesimal elasticity methods, and describe the deformation process of the soft tissue as a sequence of quasistatic equilibrium configurations, in which inertial component was not considered and the effects of geometric nonlinearities were not represented. Wilhelms-Tricarico (1995) proposed a rigorous method for modeling the soft tissue, and used it to build a three-dimensional model of the tongue. He further (1996) discussed the control strategies for tongue movement based on a neural network learning model. Payan and Perrier (1997) reported a two-dimensional biomechanical tongue model which was built using the FE method. Their model produced V-V sequences according to the Equilibrium Point Hypothesis (EPH), which is one of the common motor control methods.

In parallel to the studies for modeling the tongue tissue, a number of studies have dealt with the analysis and modeling of jaw movements during speech (Vatikiotis-Bateson and Ostry,

1995; Laboissière, Ostry, and Feldman, 1996; Ostry, Vatikiotis-Bateson, and Gribble, 1997). Vatikiotis-Bateson and Ostry (1995) reported that jaw motion is fully specified by three orientation angles and three positions. Among the six degrees of freedom, the principal components fall primarily within the midsagittal plane such as the jaw rotates downward and translates forward during opening, and the pattern is reversed during closing. The relationship between sagittal plane rotation and horizontal translation is generally linear, although instances of pure rotation and pure translation were also observed. According to these observations, Laboissière et al. (1996) proposed a control strategy for a multi-muscle system of human jaw and hyoid bone movements, which was based on the EPH.

Most of the studies on physiological models have mainly focused on the soft tissue or the rigid organs separately. In contrast, Honda et al. (1994) and Hirai et al. (1995) proposed a two-dimensional physiological model of speech organs incorporating a tongue-larynx interaction. Their model consisted of two subsystems: soft tissue and rigid body. The soft tissue of the tongue was modeled using the FE method based on MRI data from a male speaker. The rigid organs included the mandible, hyoid bone, thyroid cartilage, and cricoid cartilage. The muscles that connect those parts of the model were represented by a spring with a contractile element. The mechanical principle of the dynamic balance of forces and moments was used at the interfaces between soft and rigid structures to realize the interaction among the model components. The model reproduced the biomechanical interaction between the tongue and larynx, corresponding to MRI observations. Since the model was constructed using the FEM for the tongue and the mass-spring approach for the rigid bodies, it suffered from slow computation in achieving equilibrium between the soft tissue and rigid bodies. On the other hand, Sanguineti et al. (1998) employed a model similar to that proposed by Honda et al. (1994) and Hirai et al. (1995) to develop a control strategy based on the EPH (λ model). They showed that the dynamic effects were crucial in modeling speech-like movements because of different dynamic behaviors of the soft and bony structures.

In this study, we propose a physiological articulatory model to produce a dynamic change of the vocal tract shape. This model aims to adopt most of the advantages of the physiological models cited above, and develops new approaches for the dynamic control. In the proposed model, the tongue body was designed as a 2-cm-thick sagittal slice bounded by the midsagittal plane and two parasagittal planes, one for each side. The model geometry is based on sagittal MRI data from a male subject. Although the current model replicates only the central part of the tongue in the left-right dimension, it is referred to as a three-dimensional (3D) model in this

paper. The jaw and hyoid bone were modeled as a 3D structure on the same parasagittal planes. The outer wall of the vocal tract covering the hard palate and pharyngeal wall was constructed from the midsagittal and four parasagittal planes. All of these components were integrated to form the current model.

In this paper, Section I gives a detailed description of the geometrical data and methods for designing the model. Section II describes the computational method. Both the soft tissue and rigid structure were treated as a combination of mass-points and two kinds of connections: viscoelastic connections for the soft tissue and rigid ones for the bones. The constraints for the volume of the soft tissue and for the movement of the rigid organs are discussed. Section III develops a control strategy for the model's dynamic movements during production of vowel sequences. This section also describes a preliminary approach to the problem that occurs at the collision of the tongue with the surrounding organs. Section IV compares model simulations with kinematic articulatory data for the target speaker.

I. DESIGN OF THE ARTICULATORS

This work reports various methods used for designing and controlling the proposed physiologically-based articulatory model. The model attempts to replicate the behaviors of human speech organs. For this purpose, subject-specific customization of the model was performed by replicating the individual anatomical information that was obtained from one subject using magnetic resonance imaging (MRI) technique. The shape of the tongue was extracted based on the volumetric MRI data from a male Japanese speaker using a standard spin echo method. The anatomical data for rigid organs and the tongue muscles were extracted from different MRI data sets. Higher resolution volumetric MR images with a smaller slice thickness and a longer relaxation time (TR) were used for the identification of muscle structure.

A. Design of the Tongue Shape

The model of the tongue is designed as two soft-tissue layers bounded by three sagittal planes. Considering computational cost and practicality, the partial 3D design of the tongue is adopted to realize tongue deformation in producing vowels and consonants. The midsagittal groove of the tongue and the side airway for consonant /l/ can be represented in the model.

1. Extraction of the soft tissue

The MRI data used for modeling the tongue consisted of 15 sagittal slices. A 30 cm x 30

cm field of view for a slice of images was digitally represented by a 256 x 256 pixel matrix. The relaxation time (TR) was 500 ms and the excitation time (TE) was 20 ms. The slices were acquired in a 0.7-cm-thickness without overlap. This volumetric data set was then reconstructed using a commercial software (VoxelView) on a workstation, IRIS Indigo, where the size of the voxels was interpolated to be 0.1 cm $\times$0.1 cm $\times$0.1 cm. The desired planes of the tongue were derived from the reconstructed data. The tongue shape of a Japanese vowel [e] was chosen as the initial shape of the model. The outlines of the tongue were extracted from two sagittal slices: one is the midsagittal plane and the other is a plane 1.0 cm apart from the midsagittal on the left side. Assuming that the left and right sides of the tongue are symmetrical, the shape on the right side of the midsagittal plane was copied from the left side. Figure 1 shows the extracted outline of the tongue and the corresponding MR images. The midsagittal and parasagittal images differ from each other in the outline of the tongue, partly because the tongue dorsum forms the midline groove. The epiglottis is not included in the tongue, but represented as a part of the vocal tract. However, the volume of the epiglottis is excluded when calculating area function of the vocal tract. The outline of the tongue root in the parasagittal plane slightly exceeds the actual tongue root considering the attachment of the tongue musculature to the greater horns of the hyoid bone.

The soft tissue of the tongue consists mainly of muscles, but it also includes glands, blood vessels, fatty tissue, and surface connective tissue in addition to the mucous membrane. Each component has different physical properties, and therefore the tongue tissue is a non-homogeneous continuum. However, the tongue tissue is treated as homogeneous in the current model to apply simple physical foundations for the first approximation.

2. Modeling of the soft tissue

The soft tissue of the tongue has been modeled commonly using the FE method (Kakita, et al., 1985; Wilhelms-Tricarico, 1995). In our earlier study (Hirai, Dang, and Honda, 1995), the FEM model of the tongue was combined with a beam-muscle model of the jaw-larynx system to develop an integrated model of the speech organs. The computation of the hybrid model was slow because the achievement of the equilibrium between the soft tissue and rigid organs took considerable time. One possible solution to this problem is to model the soft tissue and rigid organs with an identical method. To the end, we designed the current model using mass-points and viscoelastic springs for both soft tissue and rigid organ.

The basic structure of the model is adopted from the fiber orientation of the genioglossus

muscle. This muscle is triangular, located parallel to the midsagittal plane of the tongue. The sagittal slice tongue model is thus represented by radiating spring connections on the midsagittal and parasagittal planes. Each plane is divided into six sections with a nearly equal interval in the anterior-posterior direction and ten sections along the tongue surface. The mesh pattern obtained from this segmentation is shown in Fig. 1. In the model, the mass-points are located in the intersections of the mesh lines, and each point is connected to the surrounding points by viscoelastic springs. The mass-points in the mid-sagittal plane also connect to the corresponding mass-points in the right and left planes by the same springs. Figure 2 shows the initial shape of the tongue model consisting of the three planes with mass-points and viscoelastic connections in different view angles, (a) for an oblique front view, and (b) for an oblique back view. The midsagittal groove is seen in the anterior portion of the tongue shown in Fig. 2 (a), and the geometry of the tongue root involving with the epiglottis is shown in Fig. 2 (b).

Energy transformation within a mass-spring system obeys the energy conservation law, as shown by

$$\frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 = \varepsilon, \qquad (1)$$

where m and k represent mass and stiffness. x is an increment of the spring length, and $\dot{x}$ is the velocity of the mass-point. ε is a constant. This equation shows that the energy transformation between the potential and kinetic components takes place when a deformation occurs. That is, the spring length must vary with the deformation in a one-dimensional system, and the energy is preserved in the length increment of the springs. Thus, the deformation can be completely restored by the potential energy preserved by the length increment of the springs if the system is lossless. However, in a multiple-dimensional mass-spring system, such a deformation does not always correspond to a length increment of the springs. For example, in a system consisting of four mass-points and four springs in a rectangular shape, there can be deformation from a rectangle to a parallelogram without any changes in the spring length. In this case, no strain forces occur in the springs during such a deformation, and thus the original shape cannot be restored. In the case of soft tissue, however, the original shape can be automatically restored from a deformation by a strain force. To realize the restoration of the model from deformation, each mass-point m_{ijk} connects to all the adjacent mass-points m_{xyz} by springs. m_{ijk} and m_{xyz} satisfy

$$m_{ijk} \bigcup m_{xyz} \subset M$$
$$m_{ijk} \neq m_{xyz} \qquad (2)$$
$$(x = \{i-1, i, i+1\}, \quad y = \{j-1, j, j+1\}, \quad z = \{k-1, k, k+1\}),$$

where M is the set of all mass-points in the model of the tongue tissue. Thus, a mass-point has many springs connecting to the adjacent ones, and the number varies from seven up to 26, depending on its location. Figure 3 shows the internal structures of the tongue model in different views: the midsagittal plane, the front view of plane Pa, and an oblique view of plane Pb. Both the solid line and the dashed line between two mass-points represent the viscoelastic springs. The solid lines show the springs in the section boundary, and the dashed lines show the springs between the diagonal mass-points. The mass-points are located in the intersections of the solid lines. With this network connection, when an external force acts on the mass-spring model, a deformation of the model results in a change in the length of the related springs, and accordingly the strain forces corresponding to the length increments are preserved in the model. Thus, the original shape can be restored from the deformation by the strain forces after the external force is removed.

The mass per unit volume is chosen to be 1 g/cm^3 for the tongue tissue, which is the same as that of water. The mass of each mass-point is determined by the unit mass and the size of each segmented unit (block), which includes the mass-point and the number of springs that share the mass-point. Note that each block in the tongue model is defined by 12 boundary springs with eight vertices indicated by solid lines in Fig. 3. Assuming that each of the eight vertices approximately possesses one eighth of the mass of the block, the mass of a mass-point is the sum of the sharing masses for the mass-point.

There are three types of models used to describe the tissue properties: the Voigt model, the Maxwell model, and the Kelvin model (Fung, 1984). All the models are composed of combinations of linear springs and a dashpot. The Voigt model is more adequate for describing solid properties, while the Maxwell model is better for fluid properties. The response of the Voigt model is of the same phase as that of the input force, while it is delayed approximately 90 degrees in the Maxwell model. The Kelvin model is a combination of these two models. When the viscous component is of the same order as the stiffness, the Voigt model and Kelvin model show no significant difference in the response phase. Therefore, we adopted the Voigt model to approximate the properties of the soft tissue of the tongue, which consists of a spring parallel to a dashpot. The mechanical parameters for the spring and the dashpot reported in the previous studies deferred widely. The stiffness ranged from 10^4-10^6

dyne/cm^2, and the viscosity from 10^5-10^7 dyne•s/cm^2 (Sakamoto and Saito, 1980). In the present model, both parameters were chosen in the same order, 1.54×10^5 dyne/cm^2 for the stiffness and 1.75×10^5 dyne•s/cm^2 for the viscosity. Since there are seven connections in a block of tongue tissue, the parameter values for a viscoelastic spring must be one seventh of the chosen values. The parameters used in this model are listed in Table 1 for the mass, viscosity and stiffness.

B. Arrangement of the Tongue Muscles

The anatomical arrangement of the major tongue muscles was determined based on a new set of high-resolution MR images obtained from a target speaker. The data-set includes 40 sagittal slices of a 0.35-cm thickness with a 0.05-cm overlap acquired during a rest position. A 25 x 25 cm field of view for each slice was digitally represented by a 256 x 256 pixel matrix. The relaxation time (TR) was 620 ms and the excitation time (TE) was 15 ms. The boundaries of muscles and bones were traced in each slice, and they are superimposed in an image so that each muscle is identified. Figure 4 shows the outlines of the muscles projected into three sagittal planes: (a) the midsagittal plane, (b) a plane 0.6 cm apart from the midsagittal, and (c) a plane 1.5 cm apart from the midsagittal plane.

Figure 4 (a) shows the genioglossus (GG), geniohyoid (GH), and mylohyoid (MH) in the midsagittal plane. Figure 4 (b), 0.6 cm from the midsagittal, depicts the superior longitudinal (SL), and inferior longitudinal (IL). Figure 4 (c), 1.5 cm apart from the midsagittal, indicates the hyoglossus (HG) and styloglossus (SG). All the tracings of muscles are combined in Fig. 4 (d). The other intrinsic muscles (transversus and verticalis) were not identifiable in the MR images. The orientation of all these tongue muscles was also examined with reference to the literature (Miyawaki, 1974; Warfel, 1993, Takemoto, 1996). The tongue muscles treated in the model are listed in Table 2.

Figure 5 shows the location for the extrinsic muscles in the model, (a) for the midsagittal plane, and (b) for the parasagittal plane. The genioglossus (GG) is the largest muscle in the tongue. It runs midsagittally in the central part of the tongue. Since this is a triangular muscle and different parts of the muscle exert a different effect on tongue deformation, it can be functionally separated into three muscle bundles: the anterior portion (GGa), middle portion (GGm), and posterior portion (GGp). The thickness of the lines represents the size of the muscle units, the thicker the line, the larger the maximum force produced. The two extrinsic muscles, hyoglossus (HG) and styloglossus (SG), are shown in the parasagittal plane. These

two muscles are designed to be symmetrical on the left and right sides.

C. Modeling of the Rigid Organs

The outlines of the rigid organs (*i.e.*, the jaw and hyoid bone, in the present work) were also traced from the MRI data for the target subject. Although the bony organs are not visualized in MR images because of the lack of water, the contours of the organs can be identified in MR images when they are surrounded by soft tissue. The data-set used for extracting the contours of the rigid organs is the same one as used in Section I A. Figure 6 (a) shows the bony framework extracted from the midsagittal plane and a parasagittal plane with a 0.7-cm interval. In the figure, the gray thick lines show the contours of the organs with reference to anatomic literature, and the dashed lines are the extracted boundary of the soft tissue. The thick dark lines show the rigid organs traced in the midsagittal plane, and the thin lines for the organs in the parasagittal plane.

Figure 6 (b) shows the model of the mandible and the hyoid bone. The right half of the mandible is drawn in the background using pale gray lines. The model of the mandible has four mass-points on each side, which are connected by five rigid beams (thick lines) to form two triangles with a shearing-beam. The shape of the triangles is invariable as long as the beam length is constant. This mandible model is combined with the tongue model at the mandibular symphysis. The temporomandibular joint is modeled to allow two types of motions: rotation and translation. The model of the hyoid bone has three segments corresponding to the body and bilateral greater horns. Each segment is modeled by two mass-points connected with a rigid beam on each side. Eight muscles (thin lines in Fig. 6) are incorporated in the model of the mandible-hyoid bone complex. The small circles indicate the attachment points of the muscles. Since the other rigid organs below the hyoid bone, such as the thyroid and cricoid cartilages, were not included in the present model, two viscoelastic springs are used to play the role of the strap muscles.

One of the problems in controlling jaw movement is the complex mechanism of opening and closing. Jaw movements in the sagittal plane involve a combination of rotation (change in orientation) and translation (change in position). The mandible rotates downward and translates forward during opening. This action is produced by the anterior belly of the digastric muscle for lowering, and by the lateral pterygoid for protrusion. The pattern is reversed during closing by relaxation of the opening muscles, contraction of the medial pterygoid, and elastic recoiling of the soft tissue. The large masticatory muscles such as the masseter and temporalis raise and

retract the jaw in chewing, but they are not active in speech movement. Although there is no one-to-one mapping between muscle actions and the kinematic degrees of freedom, the muscles that are involved in jaw movements during speech can be approximately separated into two groups: the jaw closer group and jaw opener group. These muscles are listed in Table 2.

D. Construction of the Vocal Tract Wall

The goal of the current model is to produce dynamic change in the vocal tract shape. Therefore, it is necessary to incorporate the wall of the vocal tract in the model. The surrounding organs determining a vocal tract shape comprise the lips, teeth, hard palate, soft palate (the velum), pharyngeal wall, and the laryngeal tube. At this stage, the model has no lips, and incorporates the other organs as a single rigid wall. The movements of the velum and larynx are not taken into account in the present model.

The outlines of the vocal tract wall are extracted from the MRI data described in Sec. I A. Figure 7 shows the extracted outlines of the vocal tract wall and the mandibular symphysis with the reconstructed 3D surface for the walls. In Fig. 7 (a), the thin dark lines show the walls in the midsagittal plane. The pale thick lines indicate the walls in the parasagittal plane on the right side 1.4 cm apart from the midsagittal plane, and the medium lines for the plane 0.7 cm apart from the midsagittal plane. With an assumption that the left and right sides are symmetric, a 3D surface model of the vocal tract wall and the mandibular symphysis was reconstructed using the outlines with 0.7 cm intervals in the left-right direction. Because of the geometrical complexities, it was impossible to derive an analytic function of the surface walls. For this reason, the surfaces of the tract wall and the mandibular symphysis were approximated using a number of triangular planes in the current model. The reconstructed surfaces are shown in an oblique-front view in Fig. 7 (b). The surface of the tract wall was compounded using 432 triangular planes, and 192 triangular planes for the mandibular symphysis. The piriform fossa, which has bilateral cavities and behaves as side branches in the vocal tract (Dang and Honda, 1997), was also combined in this surface model.

II. COMPUTATIONAL METHODS

The present model is a combination of the soft tissue of the tongue, the rigid vocal tract wall, the mobile bones (mandible and hyoid bone), and the related muscles. This section describes computational methods used in this model.

A. Computational Method for the Tongue, Jaw and Hyoid Bone

Since the tongue, jaw and hyoid bone are modeled as a network of mass-points and springs, the computation of the soft tissue and rigid organs can be combined into an equation system. For convenience, the term "*node*" is used hereafter to represent the mass-point and its position. Assuming that the model consists of n nodes, including the tongue, jaw and hyoid bone, we label these nodes from zero to $n-1$. The motion equation for node i ($i=0,1,...,n-1$) can be described as

$$m_i \sum_{c=1}^{cn} \ddot{\delta p}_{ic} + \sum_{c=1}^{cn} bb_{ic} \dot{\delta p}_{ic} + \sum_{c=1}^{cn} kk_{ic} \delta p_{ic} = f_i, \tag{3}$$

where p_{ic} is the viscoelastic connection of node i ($p_i=[x_i, y_i, z_i]$) and node c ($p_c=[x_c, y_c, z_c]$), where i and c meet Eq. (2). δp_{ic} is the length increment of connection p_{ic}. $\dot{\delta p}_{ic}$ and $\ddot{\delta p}_{ic}$ are the first and second derivatives of δp_{ic}. cn is the total number of the connections node i has. kk_{ic} and bb_{ic} are the stiffness and viscous component of the viscoelastic connections between node i and node c. To introduce the necessary constraints into the motion equation, as shown below, it is more straightforward to use the coordinate values than the length increment. For this reason, δp_{ic} is substituted by the coordinate value P, and its derivatives $\dot{\delta p}_{ic}$ and $\ddot{\delta p}_{ic}$ are approximated using the first order and second order differentials of P.

$$\begin{aligned}
\delta p_{ic} &= p_i(t) - p_c(t) \\
\dot{\delta p}_{ic} &= \frac{p_i(t) - p_i(t-h)}{h} - \frac{p_c(t) - p_c(t-h)}{h} \\
\ddot{\delta p}_{ic} &= \frac{p_i(t) - 2p_i(t-h) + p_i(t-2h)}{h^2} \\
&\quad - \frac{p_c(t) - 2p_c(t-h) + p_c(t-2h)}{h^2}
\end{aligned} \tag{4}$$

where t represents time and h is a time step. Substituting (4) in (3), the equation (3) is reduced to a differential equation system of P. By replacing the terms according to known and unknown variables, the equation arrives at a general form

$$(M + hB_p + h^2 K_p)P = h^2 F_p + R_p, \tag{5}$$

where M denotes a diagonal matrix consisting of the masses of all the mass-points within the model (the tongue, mandible, and hyoid bone). B_p and K_p are matrices of the viscous

components and the stiffness of all the viscoelastic springs connecting the nodes. R_p represents the sum of all known terms except the input force F_p, which depends on the previous positions of the nodes, B_p and K_p.

Since the motion equation of the model is a three-dimensional one, P in the above equations is a vector consisting of a number of sub-vectors with three elements. To reduce the 3D equations to one-dimensional simultaneous equations, the elements in P are relabeled according to the following rule. Each node i has three variables for the three dimensions. In the proposed simultaneous equations, the variables of node i are numbered as $3i$ for x-direction, $3i+1$ for y-direction, and $3i+2$ for z-direction. The new vector is referred to as X, whose element number is $N=3n$ for all nodes.

$$(M + hB + h^2K)X = h^2F + R \ , \tag{6}$$

where F is a one-dimensional vector derived from the F_p , and R is the sum of known terms. With this arrangement, the matrices and the vectors in the simultaneous equations have the following forms.

$$M = diag\{m_0, m_1, ..., m_j, ..., m_{N-1}\}, \tag{7}$$

where $j = \{3i, 3i+1, 3i+2\}$ and $m_{3i} = m_{3i+1} = m_{3i+2}$, $\quad (i = 0,1,2,...n-1)$.

$$B(t) = \begin{bmatrix} b_{0,0} & b_{0,1} & \cdots & b_{0,N-1} \\ b_{1,0} & b_{1,1} & \cdots & b_{1,N-1} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N-1,0} & b_{N-1,1} & \cdots & b_{N-1,N-1} \end{bmatrix},$$

$$K(t) = \begin{bmatrix} k_{0,0} & k_{0,1} & \cdots & k_{0,N-1} \\ k_{1,0} & k_{1,1} & \cdots & k_{1,N-1} \\ \cdots & \cdots & \cdots & \cdots \\ k_{N-1,0} & k_{N-1,1} & \cdots & k_{N-1,N-1} \end{bmatrix}, \tag{8}$$

$$X(t) = [x_0, x_1, ..., x_j, ..., x_{N-1}]^T,$$

$$F(t) = [f_0, f_1, ..., f_j, ..., f_{N-1}]^T,$$

where $k_{i,j}$ and $b_{i,j}$ are represented by the following expressions, which obey the tensor transformation law.

$$
\begin{aligned}
k_{3i+u,3i+v} &= \sum_{c=1}^{cn} kk_{ic} r_{ic}(u) r_{ic}(v), \\
k_{3c+u,3i+v} &= kk_{ic} r_{ic}(u) r_{ic}(v), \\
b_{3i+u,3i+v} &= \sum_{c=1}^{cn} bb_{ic} r_{ic}(u) r_{ic}(v), \\
b_{3c+u,3i+v} &= bb_{ic} r_{ic}(u) r_{ic}(v), \\
u &= 0,1,2; \qquad v = 0,1,2.
\end{aligned}
\tag{9}
$$

$r_{ic}(u)$ denotes the direction cosine from node i to node c: u=0 for x-direction, u=1 for y-direction, and u=2 for z-direction. Due to the symmetric properties of the viscoelastic connection between two nodes, $b_{3i+u,3c+v}$ and $k_{3i+u,3c+v}$ are equal to $b_{3c+u,3i+v}$ and $k_{3c+u,3i+v}$, respectively. Therefore, *B(t)* and *K(t)* are symmetric matrices. Although the viscoelastic components are constant over time in this model, the matrices of the viscosity and stiffness are time-varying because the direction cosine $r_{ic}(u)$ varies with time.

B. Computation of the Muscle Forces

The driving source of the model is the generalized force from muscle contraction. In this study, a commonly accepted assumption is adopted in formulating the generalized model of the muscle; a force depending on muscle length is the sum of the passive component (independent of muscle activation) and the active component (dependent on the muscle activation). Figure 8 (a) shows a diagram of the rheological model for a muscle unit, which has three parallel parts similar to Morecki (1987). Part 1 of the muscle unit consists of a linear spring k_1 and a linear dashpot b_1. Both k_1 and b_1 are defined twice as the values listed in Table 1. This part always involves the force generalization. Part 2 is a linear spring k_2, which is involved in generalizing force only when the current length l of the muscle unit is longer than its initial length l_0. The value of k_2 is determined to have a resonance frequency of about 4 Hz.

$$k_2 = 0.002k_1 \tag{10}$$

Part 3 of the muscle unit in Fig. 8 (a) models the active component, which consists of a contractile element parallel to a dashpot and then cascaded with a spring. This part plays a role when a muscle is activated. The active component of the force generalized by Part 3 can be represented using a fourth-order polynomial of the stretch ratio, $\varepsilon = (l - l_0)/l_0$, of the muscles (Morecki, 1987).

$$f(\varepsilon) = \alpha(3.42\varepsilon^4 + 0.53\varepsilon^3 - 2.2\varepsilon^2 + 0.3\varepsilon + 0.13)U \quad (11)$$

where U, a coefficient for normalization, is equal to 6.6 (Morecki, 1987). α is a gain of the active part. The relationship between the stretch ratio of the muscular unit and the generalized force is shown in Fig. 8 (b), where α =1.0. This empirical formula is valid for $-0.185 < \varepsilon < 0.5$. In our muscle model, Part 1 of the muscle unit was simplified to have one viscoelastic connection only, instead of two parallel connections with a spring and a dashpot in Morecki's model.

For a quick and smooth spread of muscle forces within the tongue model, the muscles are represented by a number of muscle units with various lengths and directions, as shown in Fig. 5. The force is generated by each unit according to the stretch ratio and the activation of the muscle, and serves as an input to the connected nodes. When a node is connected with multiple muscular units, the input force for the node is the sum of the forces of all the muscle units. The summation of the forces can be calculated by the method described in Eq. (9).

Since a whole muscle consists of many muscular units with different length and thickness, the general lumped rheological parameters of muscle tissue are not appropriate to determine the force generated by a muscle in this model. Therefore, we used two coefficients, the thickness of the muscle fiber and a gain of the force generation, in addition to the basic parameters k_1, b_1, and k_2. The first coefficient, a relative value, is used to describe the difference in thickness of the muscles based on the anatomical literature, which ranged from 0.1 to 20. This was shown in Fig. 5 where, for example, GGp is thicker than GGa. For the same muscle GGp, the muscular fiber is thinner in the layer near the surface. The other coefficient is a gain represented by α in Fig. 8 (a), which is determined by our simulation experiments. In the current model, α was equal to 6000 for all muscular units. The force generated by the active part is also dependent on the degree of muscle activation indicated by activation patterns, which is discussed in the latter part of this paper.

C. Volume Constraint for the Tongue Tissue

In the present model, the soft tissue is modeled as a network of viscoelastic springs and mass-points. With these connections alone, the model lacks the incompressible property of the tongue tissue. Therefore it must have a constraint to maintain the volume of the tongue tissue when the tongue deforms. To do this, the volume of the tongue at the initial condition as well as at each computational step was calculated, and a rule was introduced to minimize the

change in the volume. Based on the division shown in Fig. 2, the tongue tissue consists of 120 polyhedrons (blocks) with eight vertices. The whole volume of the tongue is obtained by the sum of all the sub-volumes of the blocks.

Since it is not guaranteed that any four adjacent vertices of a block in the model are coplanar, there is no analytic expression to calculate the volume of such a block. The only way to obtain the volume of such a block is to divide it into several tetrahedrons. Basically, there are two different ways to uniquely divide a block with eight vertices into five tetrahedrons, regardless of its shape. The volumes obtained from these two different ways of divisions are not consistent in most of the cases of model computation. To arrive at a consistent solution, the average value of the volumes from the two ways of divisions is used as the estimated volume for each block. Thus, the volume (V_j) of block j is

$$V_j = \frac{1}{2}\sum_{s=1}^{2}\sum_{i=1}^{5} v_{si},$$
$$v_{si} = \begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_3 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}, \tag{12}$$

where $\{x_p, y_p, z_p\}(p = 1,...,4)$ are the coordinate values of vertex p in tetrahedron i. v_{si} is the volume of tetrahedron i obtained from the dividing method s (s=1, or 2). The whole volume $\hat{V}$ of the tongue tissue is the sum of the volumes of all the blocks.

$$\hat{V} = \sum_{j=1}^{dn} V_j, \qquad dn = 120, \tag{13}$$

where dn is the total number of the blocks in the tongue, which is 120 in the present model.

The square of the difference between the current volume and the initial volume was defined as an observation error. The control error is the sum of the observation errors for both the whole body of the tongue and all of the individual blocks.

$$e = (\hat{V} - \hat{V}_0)^2 + \sum_{j=1}^{dn} (V_j - V_{j0})^2, \tag{14}$$

where $\hat{V}_0$ is the initial volume for the whole body of the tongue and V_{j0} for block j. The control error is minimized by setting the partial derivative of the observation error in (14) with respect

to x_i to zero.

$$\frac{\partial e}{\partial x_i} = \frac{\partial (\hat{V} - \hat{V}_0)^2}{\partial x_i} + \sum_{j=1}^{dn} \frac{\partial (V_j - V_{j0})^2}{\partial x_i} = 0, \qquad i = 0,1,...,N-1 \tag{15}$$

Expression (15) consists of N simultaneous equations. Node positions solved from these equations must satisfy the constraint that the volume changes of the tongue are minimum at each computational step.

Since expressions (5) and (15) are two independent systems of simultaneous equations, they usually give discrepant solutions for node positions. In order to avoid this situation, a tradeoff is made between the motion equations and the constraint equations. This tradeoff allows non-zero residues for the motion equations, and it then minimizes the sum of the squared residues and the volume errors by setting the partial derivative of the summed error with respect to x_i to zero. With this tradeoff, the computational equation is

$$\frac{\partial \left| (M + Bh + Kh^2)X - Fh^2 - R \right|^2}{\partial x_i} + \alpha_1 \frac{\partial (\hat{V} - \hat{V}_0)^2}{\partial x_i} + \alpha_2 \sum_{j=1}^{dn} \frac{\partial (V_j - V_{j0})^2}{\partial x_i} = 0, \qquad i = 0,1,...,N-1, \tag{16}$$

where α_1 and α_2 are the coefficients for adjusting the tolerance of the changes in the volumes. The volume tolerance was adjusted to be less than 2% in the present study.

D. Constraints for Jaw and Hyoid Bone Movements

Previous studies have revealed that jaw movement during speech is not a pure joint rotation but a combination of rotation and translation (Ostry and Munhall, 1994). The hyoid bone also shows rotation and translation since it has no joint connection to the surrounding structure. Therefore, this rigid structure in the model can be assumed to be a floating body, and its position and orientation are determined by the sum of active and passive elastic forces applied.

Although the movements of the jaw and hyoid bone should be treated as dynamic rotation and translation components of a rigid body, it is difficult to combine these components and tongue deformation within a linear motion equation system. If assuming that the jaw and hyoid bone are composed of several mass-points with rigid beams, however, we can model

them as a part of the mass-spring system by adopting a constraint that maintains the length of the beams and limits the moment of the beams. In the present model, the length constraint was implemented by setting a very large stiffness and viscosity, whose values were approximately ten thousand times greater than those used for the soft tissue.

In the model, the mandible is driven to rotate by a torque at the rotation center of the condyle. The torque is calculated by the total forces on the mandible and jaw orientation angle. The jaw rotation affects the tongue body via the attachments on the mandible, and also reflects on the hyoid bone through the tongue tissue and the other muscles. For the hyoid bone, the forces acting on the attachments are redistributed via the moment about the center of the hyoid bone.

To simulate the translation of the jaw, we adopted an assumption that the movement of the anatomical center of rotation of the condyle follows a curved path corresponding to the concave articular groove of the temporomandibular joint (Laboissière et al.,1996). The curved path is approximated using a third-order polynomial:

$$y = y_0 - 2.5(x - x_0)^3 - 4(x - x_0)^2, \tag{17}$$

where x and y are the horizontal and vertical positions of the rotation center of the condyle. x_0 and y_0 are the initial positions of the condyle. A consequence of this simplification is that the vertical position of the anatomical center of rotation of the jaw is wholly dependent on its horizontal position (Ostry and Munhall 1994).

Figure 9 shows the curved path of the temporomandibular joint, given in (17). One can see that this function shows a realistic path in the anterior portion, which serves as a railway for the condyle. The function also gives an abruptly declined slope that can avoid the jaw going back too far during closing. The constraint of the curved path for the condyle is given by minimizing the distance of the condyle to the curved path.

$$\begin{aligned} e &= [y - y_0 + 2.5(x - x_0)^3 + 4(x - x_0)^2]^2 \\ \frac{\partial e}{\partial x} &= 0 \\ \frac{\partial e}{\partial y} &= 0 \end{aligned} \tag{18}$$

This constraint equation is also combined in equation (16).

When the condyle is at a given position along the articular groove, the total force exerted

by the muscles on the jaw can be decomposed into the component on the tangential line (*i.e.*, the slope) of the articular path at the contact point and another component perpendicular to that line. The latter component will be counteracted by the joint reaction force while the former one will be responsible for jaw movements. The tangential component is redistributed into the horizontal and vertical axes to drive the jaw. Figure 9 shows the force redistribution at the contact point of the condyle with the curved path. When x-direction force f_x and y-direction force f_y act on the condyle, both the forces respectively decompose into two force components parallel to the slope of the curved path and perpendicular to the slope. The dark lines show the components derived from f_x, and the gray lines indicate the components from f_y. The line with circles shows the sum of the forces perpendicular to the slope, which does not contribute to jaw movement. The lines with a V-shape arrow show the forces parallel to the slope. The sum of the forces that contributes to jaw movement is implemented in the model by decomposing it into x-direction and y-direction components, shown in the lines with a white arrow. Due to the curved path of the articular groove, the initial input f_x and f_y are redistributed as effective forces $f_{x'}$ and $f_{y'}$ in the model. The effective force $f_{y'}$ in y-direction has an opposite direction to the input f_y in this example.

There is another factor to be considered in the mechanism of jaw movement. As shown in Fig. 6, the distance between the posterior edge of the condyle neck and the tympanic bone is small. When the jaw opens wide by rotation alone, the condyle neck makes contact with the front edge of the tympanic bone. In actuality, jaw rotation is limited by the soft tissue in front of the tympanic bone, and the reaction force from the tissue produces a force to translate the mandible forward. In the model, the reaction force is generated when the condyle reaches the boundary, and it is decomposed into a forward component and a downward component in the articular groove according to the mechanism shown in Fig. 9.

III. DYNAMIC ARTICULATORY CONTROL

The physiological model of the speech organs presupposes movement control signals by muscle contraction patterns according to the physiological process of the motor system. Kakita et al., (1985) used electromyography (EMG) data from the extrinsic tongue muscles (Baer et al., 1988) to drive the 3D tongue model and generated formant patterns of English vowels. Maeda and Honda (1994) used the same EMG data to generate the formant patterns using Maeda's

articulatory model based on the idea of the simple mapping function between muscle contraction pattern and vowel formant pattern. Honda et al. (1994) and Hirai et al. (1995) also used EMG signals to excite the 2D model of the tongue and larynx, and generated Japanese vowel formants and a natural pattern of intrinsic vowel F0. These attempts only used a limited set of available EMG data and had to apply arbitrary muscle contraction patterns for the other muscles.

A different approach has been developed for obtaining a control strategy without empirical EMG data. Assuming that speech control is target-based, Laboissière (1996) and Perrier et al. (1996) studied the control of dynamical models of the articulators based on the Equilibrium Point Hypothesis (EPH), which argues that limb movements are produced by centrally specified shifts of the mechanical equilibrium of the peripheral motor system. Payan and Perrier (1997) and Sanguineti et al. (1998) also used the EPH control strategy in their two-dimensional biomechanical tongue model.

The EPH method is basically a plausible approach to control dynamic movements of the tongue and other articulators, since the articulatory system is a network of muscles and floating rigid bodies that exhibits relatively slow movement. The movement of such a system can be estimated reasonably well only by the balance of all the forces applied on the network. However, this method requires a muscle length parameter that is empirically difficult to obtain. A practical method to drive a physiological model is to speculate on muscle activation patterns that resemble the EMG signals. In this study, a new control strategy for a physiological articulatory model was developed based on the assumption that the muscle activation patterns are dependent on the current status and the target status of the articulators.

A. Tongue Deformation by Muscle Contraction

The first approach towards developing a control strategy is to examine the effect of individual muscle contraction on tongue deformation. Figure 10 shows the results of tongue deformation by contraction of the extrinsic tongue muscles. The tongue body moves upward and forward by the contraction of the GGp, and it moves downward and backward by the HG, as shown in Fig. 10 (a) and (d). As shown in Fig 10 (b) and (c), the tongue body moves backward and upward by the SG, and moves downward and forward with a deepened-groove by the GGa. The arrows show the direction of tongue movements due to the contraction of each muscle. These simulation results are basically consistent with the observations in EMG experiments (Baer et al., 1988; Honda et al., 1992; Dang and Honda, 1997).

Kusakawa et al, (1993) and Honda (1996) described the function of the four major extrinsic muscles as two pairs of antagonistic muscles. The GGp and HG form the major antagonistic pair, and the SG and GGa form another pair. Based on the fact that the two antagonistic pairs are approximately orthogonal, Kusakawa et al, (1993) computed the equilibrium of the four extrinsic muscle forces (represented by normalized EMG signals) in the space consisting of the two antagonistic pairs in order to derive articulatory trajectories of the combined muscle forces. Based on the trajectory pattern, Honda (1996) explained a simple relationship of vowel representations in the articulatory space and the auditory space. In this study, we employed a similar idea to establish a relationship between articulation and muscle activities.

B. Generation of Muscle Activation Signals

The new control method for articulatory movements is based on the assumption that muscle activation patterns are determined by the geometrical distance between the current position and the next articulatory target. Since the activities of the major extrinsic muscles can determine tongue shape for vowels, this study employed four extrinsic muscles of GGp, HG, SG, and GGa for the tongue, and the jaw opener and closer muscles for the jaw.

The key issue in developing a target-based control strategy is to determine a tongue position at an arbitrary moment in the geometrical space. To do so, the *tongue position* was defined by the average position of five midsagittal node points on the tongue surface from the tongue tip to the back of the tongue. The movements of the *tongue position* were computed by independently contracting each muscle with a unit level of activation signal for a duration of 200 ms. Figure 11 shows the changes in *tongue position* produced by the unit contraction of the four extrinsic muscles. In the figure, the thick arrow represents a vector of displacement of the *tongue position* from the initial position. These muscle vectors form a space, referred to as a muscle workspace, which can be directly mapped on the geometrical space.

This relationship suggests that muscle activation patterns for a tongue movement towards a target position can be derived from the current tongue position and the next articulatory target. Supposing that the tongue is located in the current position *Pc* and moves forward to the target *Tg* in Fig. 11, the line from *Pc* to *Tg* forms a vector, which is referred to as an articulatory vector. When the articulatory vector is mapped onto the muscle workspace, a set of projections are obtained for each muscle vector. Although the obtained muscle projections may be positive or negative for each muscle vector, the positive projection alone provides an

activating signal whose magnitude is proportional to the projection's amplitude. Figure 11 shows that the SG and HG are the active muscles at the current computational step. When the activation signals are computed at each computational step and drive the tongue to move to a new position, a trajectory of *tongue position* is obtained as indicated by the thin gray arrow in Fig. 11. Since the workspace is not orthogonal, the direction of the initial vector formed by f_{sg} and f_{hg} in Fig. 11 does not coincide with the direction of the initial articulatory vector.

A similar muscle workspace was constructed for the jaw muscles. The activation signals for the opening and closing muscle groups are generated in the same way using a muscle workspace for the jaw muscles.

Strictly speaking, the above process faces the inverse problem of deriving muscle activity patterns from tongue positions. There can be many combinations of forces among the muscles because the same tongue position can be produced by different levels of co-contraction between the agonist and antagonist muscles. However, no attempt was made to solve this problem in this work.

C. Collision of the Tongue and the Outer Wall

In speech articulation, the tongue makes contact with the teeth, hard palate, and mandible as it moves. The tongue tip often collides with the outer wall when producing consonants such as /t/ and /l/. The lateral surfaces of the tongue body contact the hard palate to form a narrow channel of the vocal tract in production of vowel /i/. Thus, the contact between the tongue and the outer wall is one of the critical factors in achieving accurate and stable control of the tongue.

Since the outer wall was not modeled as an analytic function, the collision of the tongue on the wall cannot be combined with the motion equations of the model. Alternatively, a method was developed to compute tongue deformation when a rigid wall is introduced in the movement path of the nodes. Figure 12 shows a diagram for explaining this method. Suppose that the points P_{01}, P_{11}, and P_{21} represent the positions of three nodes m_0, m_1, and m_2 on the tongue surface at time 1. When a certain force is applied, the three nodes would move to P_{02}, P_{12}, and P_{22} at time 2 if there were no vocal tract wall. The dashed lines with an arrow show the pathways of the nodes. Since there is a wall in the pathway of the nodes, node m_0 will hit the wall at Ps and receives a reaction force. Then, the node should arrive at an equilibrium position P'_{02}, where Ps is the intersection of the trajectory of m_0 and the wall.

The equilibrium position of a node contacting the wall can be examined by the forces on

node m_0 at the collision on the outer wall. The total energy of m_0 at the collision can be approximately represented by the potential energy from Ps to P_{02}. The momentum ($m_0 v_0$) of node m_0 in a given direction is proportional to the component of the vector from Ps to P_{02} in the same direction. When the node hits the outer wall, the reaction force of the wall on m_0 is opposite of the direction to the velocity of the node v_0 and proportional to the velocity amplitude. The tangential component of the force on m_0 can be reasonably represented by the projection of the vector of Ps to P_{02} on the wall surface of a triangular plane node m_0 hit. Assuming that the plane has homogeneous properties in all directions, displacement of P_{02} on the triangular plane in each direction is proportional to the component of the momentum, *i.e.*, the vector of Ps to P_{02}. Therefore, the equilibrium position can be approximated using the projection of P_{02} on the triangular plane, shown by P'_{02} in Fig. 12 (b).

Since the vocal tract wall consists of triangular planes, a problem occurs when the collision takes place near the edge of a triangle. When the projection of P_{02} falls out of the triangular plane, it is difficult to predict the equilibrium point of node m_0. In this case, the equilibrium position is forced to move to the point at the edge of the triangular plane where the projection of the vector of Ps to P_{02} intersects with.

This collision also changes the length of the springs connected with the node, as shown in Fig. 12. For example, the spring between P_{12} and P_{02} is shortened to the one between P_{12} and P'_{02} due to the collision. This effect induces additional forces for the nodes adjacent to this node. These forces were calculated according to the length increments of the concerned springs in comparison with the length in the case without the outer wall. The forces are taken into account in the computation as an input for the next step.

IV. COMPARISONS OF MODEL SIMULATION AND OBSERVED DATA

The whole image of the proposed model is shown in Fig. 13, which consists of the tongue, jaw, hyoid bone, and the outer wall. The details of the model components are shown in Table 3. This model can demonstrate movements of the tongue and jaw according to a given sequence of articulatory targets, and a dynamic movement of the vocal tract is obtained along with a realistic collision of the tongue on the outer wall. In this section, the dynamic behavior of the articulatory model is compared with observed materials.

A. Observation of Articulator Movements

Dynamic articulatory data were obtained by an experiment using the University of Wisconsin X-ray Microbeam System. The X-ray microbeam system tracks 10-12 pellets (metal markers) placed on the articulators at an aggregate rate of up to 800 samples per second (Westbury, 1991). Figure 14 shows a diagram of the pellet placement for observing articulatory movements using the microbeam system. In total, twelve midsagittal pellets were used for the experiment. Among them, six pellets were placed on the tongue surface where four pellets, T1 through T4, were glued on the tongue surface. T5 was a pellet hanging from T4 in the back of the tongue, and T6 was a pellet dropped in the vallecula. With this placement of the pellets, the tongue contour can be obtained by interpolating the six pellets. The utterances used were vowel sequences with three vowels /a/, /i/, and /u/. The same target speaker served as a subject for this experiment.

In the phonetic literature, the tongue position for a vowel is described by the location of the highest point of the tongue dorsum. However, this representation is not compatible with the pellet data obtained from the microbeam system. Therefore, in the current stage, we used average values over three tongue pellets of T1, T2, and T3 to describe the tongue position from the microbeam data. Figure 15 shows trajectories of the tongue position during the utterances: (a) for the vowel sequences /aia/, /aua/, and /iui/, and (b) for the vowel sequences /iai/, /uau/, and /uiu/.

B. Comparison of Simulations and Observations

A set of articulatory targets were first obtained by extracting average values of the tongue position and mandibular incisor position from the microbeam data. In the simulation, the duration of each vowel in the sequences was 0.3 seconds. An example of the parameters for producing vocal tract shapes of /aia/ and /iai/ is shown in Table 4. The computed trajectories of tongue position are shown in Fig. 16, (a) for the vowel sequences /aia/, /aua/, and /iui/, and (b) for the vowel sequences /iai/, /uau/, and /uiu/.

Comparing the results shown in Fig. 15 and Fig. 16, the computed trajectories show a good agreement with the observed ones for the same vowel sequence. The similarity between computed and observed trajectory patterns is found across different vowel sequences. For instance, the tongue moves considerably more in the anterior-posterior direction in the sequence /iui/ than in /uiu/ in Fig. 15. It is interesting to find that tongue trajectories for the vowel sequences /aua/ and /uau/ show a clockwise movement. The upward and downward

paths separate more for /uau/ than for /aua/. Our model demonstrates this phenomenon well as shown in Fig. 16.

The comparison between jaw movements in the model and the observations suggested that the trajectory of the Man_I pellet roughly coincides with the movement of the mandibular symphysis in the simulation. Compared to the initial jaw position, the jaw rotates by 4.28 degrees and translates by 0.4 cm to open, and it rotates by -2.75 degrees to close with no translation.

Figure 17 shows the trajectories of ten pellets on the speech organs and eleven points in the tongue model in the simulations for the utterance /iai/. It is evident in the figure that the trajectory patterns are similar in the simulation and the observation.

C. Evaluation of the Muscle Activation Patterns

The muscle activation patterns were obtained using the proposed control strategy by stepwise computation toward the articulatory targets. Figure 18 shows an example of the results. Fig. 18 (a) plots the muscle workspace and tongue trajectory. There are two articulatory targets for /i/ and /a/, shown by the open circles. The arrows in the trajectory show the direction of the articulatory vector for the tongue movement. The gray lines show the muscle vectors in the muscle workspace.

Figure 18 (b) shows the waveforms of the activation signal for the four extrinsic muscles and for the opener and closer groups of the jaw muscles. It is interesting to find that the muscle activation signals resemble the EMG signals observed in the physiological experiments (Baer et al, 1988; Dang and Honda, 1997). The fact that the antagonistic muscles show a reciprocal pattern suggests that the proposed method, which is based on articulatory targets, can be used in place of the method using EMG signals. Since the muscle activation signals can be uniquely obtained by a given target sequence, this control strategy offers a practical method for driving a physiological articulatory model.

V. CONCLUSION

A physiological articulatory model was developed that consists of the tongue tissue, mobile jaw and hyoid bone, and the outer wall of the vocal tract. The MR images obtained from one Japanese male speaker served as geometrical data for the present model. The

articulatory data were measured using the microbeam system for the same speaker, and were used to confirm the model performance.

Mass-points and viscoelastic springs were employed in modeling the soft tissue of the tongue. This modeling method showed some advantages over the other physiological models using the finite element method (FEM). First, the system consisting of mass-points and viscoelastic springs can easily represent a large deformation of a soft tissue continuum. Second, the soft tissue and rigid organs can be integrated in the motion equation system by using two kinds of connections between the mass-points: viscoelastic ones for the soft tissue and rigid ones for the bones. The computing time of our model is about 50 times of real time using the Sun Workstation Ultra-30.

A practical control strategy was developed to produce dynamic articulation for vowel production. This strategy is a reiterative procedure; it generates muscle activation signals based on a current position and an articulatory target via the operation in the muscle workspace, and it drives the tongue toward the articulatory target. This procedure results in time-varying activation patterns for the muscles, which were similar to the EMG signal patterns.

There are several problems remaining in the proposed model for producing speech sounds. In the current stage, we only use one point to control tongue movement. The production of consonants requires control on tongue movement at multiple points, such as the tongue tip and tongue dorsum. The next study will expand the single muscle workspace of the tongue position to a multiple muscle workspace, which may consist of three single muscle-workspaces for the tongue tip, tongue dorsum and tongue root. Another issue concerns the inverse problem in generating muscle activation signals from positional data. This problem is due to the co-contraction of agonist and antagonist muscles, which was not handled in the present approach and will remain for further study.

REFERENCE

Baer, T., Alfonso, J., and Honda, K. (1988). "Electromyography of the tongue muscle during vowels in /epvp/ environment," Ann. Bull. R. I. L. P., Univ. Tokyo, 7, 7-18.

Coker, C. H. (1976). "A model of articulatory dynamics and control," Proc. IEEE 64, 452-460.

Dang, J. and Honda, K. (1997a). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. 101, 456-465.

Dang, J. and Honda, K. (1997b). "Correspondence between three-dimensional deformation and EMG signals of the tongue," Proc. of ASJ spring meeting, 241-242.

Dang, J., Honda, K. and Suzuki, H. (1994). "Morphological and acoustical analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am. 96, 2088-2100.

Fung, Y.C. (1984). *Biomechanics - Mechanical properties of living tissue*, Springer-Verlag, (2nd Edition)

Hashimoto, K. And Suga, S. (1986). "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue," J. Acoust. Soc. Jpn. (E), 7, 39-46.

Hence, W.L., (1967). "Preliminaries to speech synthesis based on an articulatory model," Proc. Of the 1967 IEEE Boston Speech Conference (IEEE, New York). 170-177.

Hirai, H., Dang, J., and Honda K. (1995)" A physiological model of speech organs incorporating tongue-larynx interaction," J. Acoust. Soc. Jpn, 52, 12, 918-928. (in Japanese)

Honda, K. (1996). "An EMG analysis of sequential control cycles of articulatory activity during /epvp/ utterances," J. Phonetics, 24, 39-52.

Honda, K., Hirai, H., and Dang, J. (1994) "A physiological model of speech organs and the implications of the tongue-larynx interaction," Proc. ICSLP 94, 175-178, Yokohama.

Honda, K., Kusakawa, A., and Kakita, Y. (1992). "An EMG analysis of sequential control cycles of articulatory activity during /epvp/ utterances," J. Phonetics, 20, 53-63.

Kakita, Y. and Fujimura, O. (1977). "Computational of tongue: a revised version," J. Acoust. Soc. Am. 62, S15(A).

Kakita, Y., Fujimura, O., and Honda, K. (1985). "Computational of mapping from the muscular contraction pattern to formant pattern in vowel space," In *Phonetic Linguistics*, edited by A. L. Fromkin, (Academic, New York).

Kiritani, S., Miyawaki, K., Fujimura, O., and Miller, J. (1976). "A computational model of the tongue," Ann. Bull. Res. Inst. Logoped. Phoniatrics Univ. Tokyo, 10, 243-251.

Kusakawa, N., Honda, K., and Kakita, Y. (1993). "Construction of articulatory trajectory in the space of tongue muscle contraction force," *ATR Technical Report*, TR-A-0171. (in Japanese).

Laboissière, R., Ostry, D., and Feldman, A. (1996). "The control of multi-muscle system: human jaw and hyoid movement," Biol. Cybern., 74, 373-384.

Maeda, S. (1990). "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp.131-149.

Mermelstein, P. (1973). "A digital simulation method of the vocal tract system," J. Appl. Physiol. 54, 1681-1686.

Miyawaki, K. (1974). "A study of the musculature of the human tongue," Ann. Bull. Res. Inst. Logoped. Phoniatrics, Univ. Tokyo, 8, 23-50.

Morecki, A. (1987). "Modeling, mechanical description, measurements and control of the selected animal and human body manipulation and locomotion movement," Biomechanics of Engineering - modeling, simulation, control, Edited by Morecki, Springer-Verlag, New York.

Ostry, D. and Munhall, K. (1994). "Control of the jaw orientation and position in mastication and speech," J. Neurophysiol, 71, 1515-1532.

Ostry, D, Vatikiotis-Bateson, E., and Gribble, P. (1997). "An examination of the degrees of freedom of human jaw motion in speech and mastication," J. Speech Hearing Research, 40, 1341-1351.

Payan, Y. And Perrier, P. (1997). "Synthesis of V-V sequences with a 2D biomechanical tongue shape in vowel production," Speech Commin., 22, 185-206.

Perkell, J. (1974). "A physiological-oriented model of tongue activity in speech production," Ph. D. Thesis, MIT.

Perrier, P., Ostry, D., and Laboissière, R. (1996). "The equilibrium point hypothesis and its application to speech motor control," J. Speech Hearing Research, 39, 365-378.

Saltzman, E. L. and Munhall, K. G. (1989). " A dynamic approach to gesture patterning in speech production," Ecol. Psychol. 1, 333-382.

Sakamoto, T. and Saito, Y. (1980). *Bionics and ME - From the basic to measurement control*, Tokyo Electric Machinery University.

Sanguineti, V., Laboissière, J., and Ostry, D. (1998). "A dynamic biomechanical model for

neural control of speech production," J. Acoust. Soc. Am. 103, 1615-1627.

Sondhi, M. and Schroeter, J. (1987). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE, Tans. Acoust. Speech Signal Process. ASSP-37, 955-967.

Takemoto, H. (1996). "Comparative anatomy of the tongue muscles of humans and chimpanzees," Abstract from the 16th Congress of International Primatological Society, August 11-16, Madison, WI, pp.418.

Vatikiotis-Bateson, E. and Ostry, D. (1995). "An analysis of the dimensionality of jaw motion in speech," J. Phonetics, 23, 101-117.

Warfel, J. (1993). *The head, neck, and trunk*, Led & Febiger, Philadelphia and London

Westbury, J. R. (1991). "The significance and measurement of head position during speech production experiments using the x-ray microbeam system," J. Acoust. Soc. Am. 89, 1782-1791.

Wilhelms-Tricarico, R. (1995). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," J. Acoust. Soc. Am. 97, 3805-3898.

Wilhelms-Tricarico, R. (1996). "A biomechanical and physiological-based vocal tract model and its control," J. Phonetics, 24, 23-48.

Table 1 Mass(M), viscosity (B), and stiffness (K) used in this model

M	1.0	g/cm^3
B	25000	$dyne \cdot s/cm^2$
K	22000	$dyne/cm^2$

Table 2 Organs and Muscles in the Model

Organs	Tongue, Hyiod bone, Mandible
Tongue muscles	Genioglossus (GGa, GGm, GGp), Styloglossus (SG), Hyoglossus (HG), Longitudinalis (SL, IL), Transversus, Verticalis, geniohyoid (GH), mylohyoid (MH).
Jaw muscles	Digastric [o], Lateral Pterygoid [o,c], Medial Pterygoid [c], Masseter, Stylohyoid, Sternohyoid, Temporalis, Geniohyoid, Mylohyoid

[o] The major muscles in the opener group;

[c] The major muscles in the closer group;

Table 3 Construction of the present model

	Tongue-body	Rigid-organs	Muscles	Tract-wall	Mandible
Nodes	231	12	150	-	-
Connections	1946	12	326	-	-
Sub-planes	-	-	-	432	196

Table 4 Parameters used in computing vocal tract shape for /iai/ and /aia/ (TP-x and TP-y are x and y positions of the tongue, JP-x and JP-y for the jaw)

/iai/	Time (s)	0.3	0.3	0.3
	TP-x (cm)	0.4	1.9	0.4
	TP-y (cm)	6.0	5.1	6.0
	JP-x (cm)	-2.9	-2.7	-2.9
	JP-y (cm)	3.5	3.3	3.5
/aia/	Time (s)	0.3	0.3	0.3
	TP-x (cm)	1.9	0.4	1.9
	TP-y (cm)	5.1	6.0	5.1
	JP-x (cm)	-2.7	-2.9	-2.7
	JP-y (cm)	3.3	3.5	3.3

Figure Captions

Figure 1 Extraction and modeling of tongue body based on volumetric MR images in (a) midsagittal plane and (b) parasagittal plane 1 cm apart from the midsagittal plane.

Figure 2 Initial shape of the tongue model: (a) oblique front view, and (b) oblique back view. Tongue model is a thick slice of the midline body, consisting of three vertical planes with mass-points and springs.

Figure 3 Internal structure of the tongue model in three-plane view. Mass-points are located in intersections of the springs shown by solid lines. Dashed lines are the springs connecting between diagonal mass-points.

Figure 4 Extraction of tongue muscles and outline of the vocal tract: (a) midsagittal plane, (b) parasagittal plane (0.6 cm), (c) parasagittal plane (1.5 cm), and (d) superimposed view of the extracted outlines.

Figure 5 Structure of extrinsic muscles of tongue model: (a) three bundles of genioglossus muscles (GGa, GGm and GGp) on midsagittal plane and (b) hyoglossus (HG) and styloglossus (SG) on parasagittal plane.

Figure 6 Modeling of the rigid organs based on MR images: (a) extracted framework of bony organs, and (b) model of the mandible and hyoid bone with related muscles (structure of the concerned muscles in reference to anatomical literature).

Figure 7 Modeling of the vocal tract wall: (a) extracted outlines of vocal tract wall based on MR images, and (b) reconstructed surface of vocal tract walls assuming symmetric left and right sides (dimensions in cm).

Figure 8 Muscle modeling: (a) a general model of muscle: k and b are stiffness and dashpot, E is contractile element. (b) generalized force varies with stretch ratio ε. α is a coefficient of the active part, depending on muscle thickness and activation level.

Figure 9 Model of the articular groove of the temporomandibular joint and forces applied on the condyle. Force is redistributed at contact point of the condyle with the path. Dark lines shows forces related to x-direction and gray lines for y-direction. *fx* and *fy* are initial inputs. *fx'* and *fy'* are redistributed forces.

Figure 10 Tongue deformations by contracting extrinsic tongue muscles: (a) tongue dorsum

advances and rises by GGp, (b) rises and retracts by SG, (c) lowers and grooves by GGa, and (d) retracts and lowers by HG. Cross arrow indicates the direction of tongue movements by indicated muscles.

Figure 11 Muscle workspace of the tongue showing a method to compute muscle contraction force. ○: initial position of the tongue; *Pc*: current position of the tongue, *Tg*: articulatory target. *fsg* and *fhg* are positive projections of vector of *Pc* to *Tg* in muscle workspace. Gray line shows trajectory from *Pc* to *Tg*, and gray arrow indicates the direction at current step.

Figure 12 Modeling of tongue collision on vocal tract wall: (a) deformation during collision and (b) new position of the node on the wall. *Pij*: Position of node *i* at time *j*.

Figure 13 An oblique view of three-dimensional model of the speech organs. All dimensions are in cm.

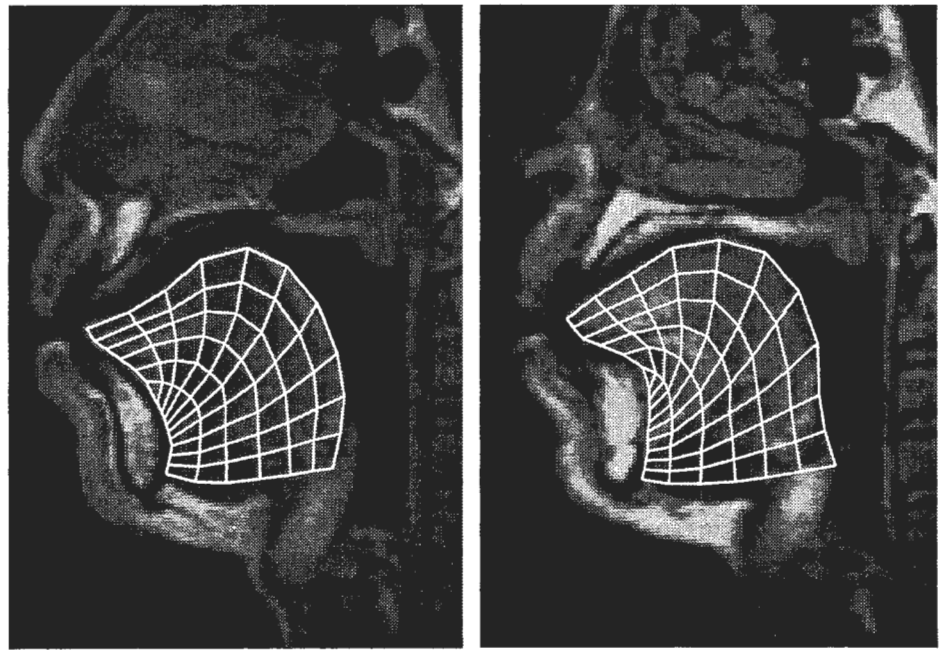
Figure 14 Pellet placements for x-ray microbeam experiments.

Figure 15 Observed trajectories of the tongue position (average of T1-T3) during three types of vowel sequences.

Figure 16 Generated trajectories of the tongue position (average of the same region as that in Fig. 15) for the same vowel sequences.

Figure 17 Trajectories of surface points of the speech organs during /iai/ in observations and simulation: (a) movements of pellets (see Fig. 14 for details), and (b) movements of model's surface nodes.

Figure 18 Model control using an articulatory target sequence: (a) articulatory trajectory of tongue position in tongue muscle workspace, and (b) generated muscle activation signals, where the signals of jaw opener and closer groups were obtained similarly from jaw muscle workspace.



(a) (b)

Figure 1 Extraction and modeling of tongue body based on volumetric MR images in (a) midsagittal plane and (b) parasagittal plane 1 cm apart from the midsagittal plane.

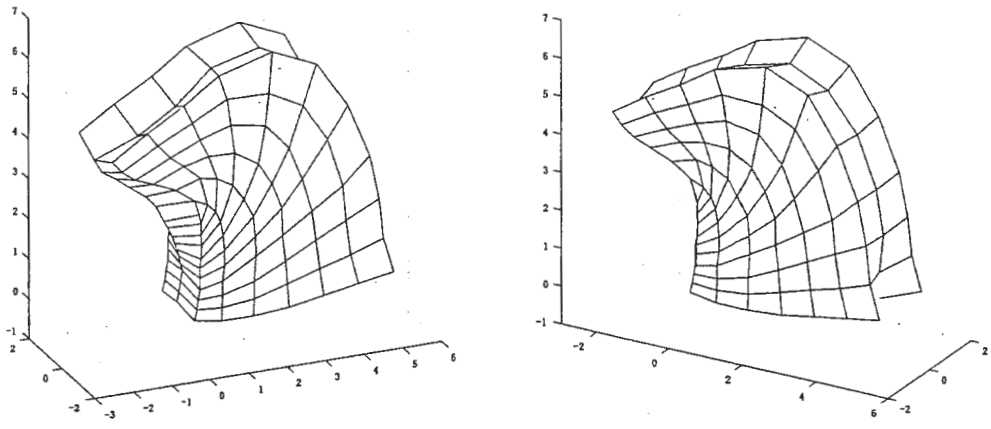


Figure 2 Initial shape of the tongue model: (a) oblique front view, and (b) oblique back view. Tongue model is a thick slice of the midline body, consisting of three vertical planes with mass-points and springs.

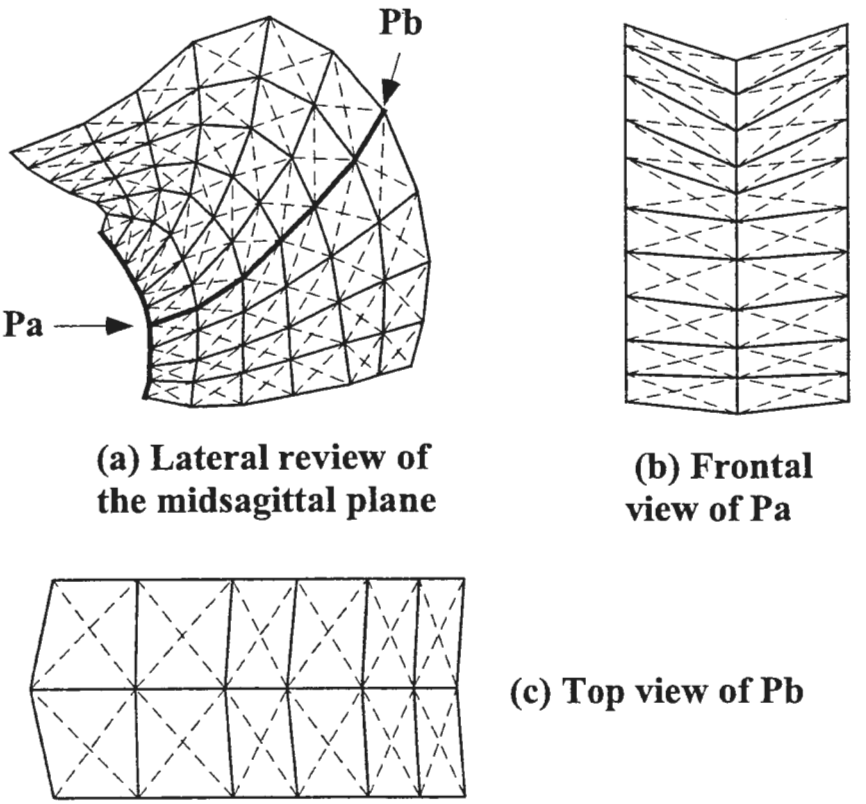


Figure 3 Internal structure of the tongue model in three-plane view. Mass-points are located in intersections of the springs shown by solid lines. Dashed lines are the springs connecting between diagonal mass-points.

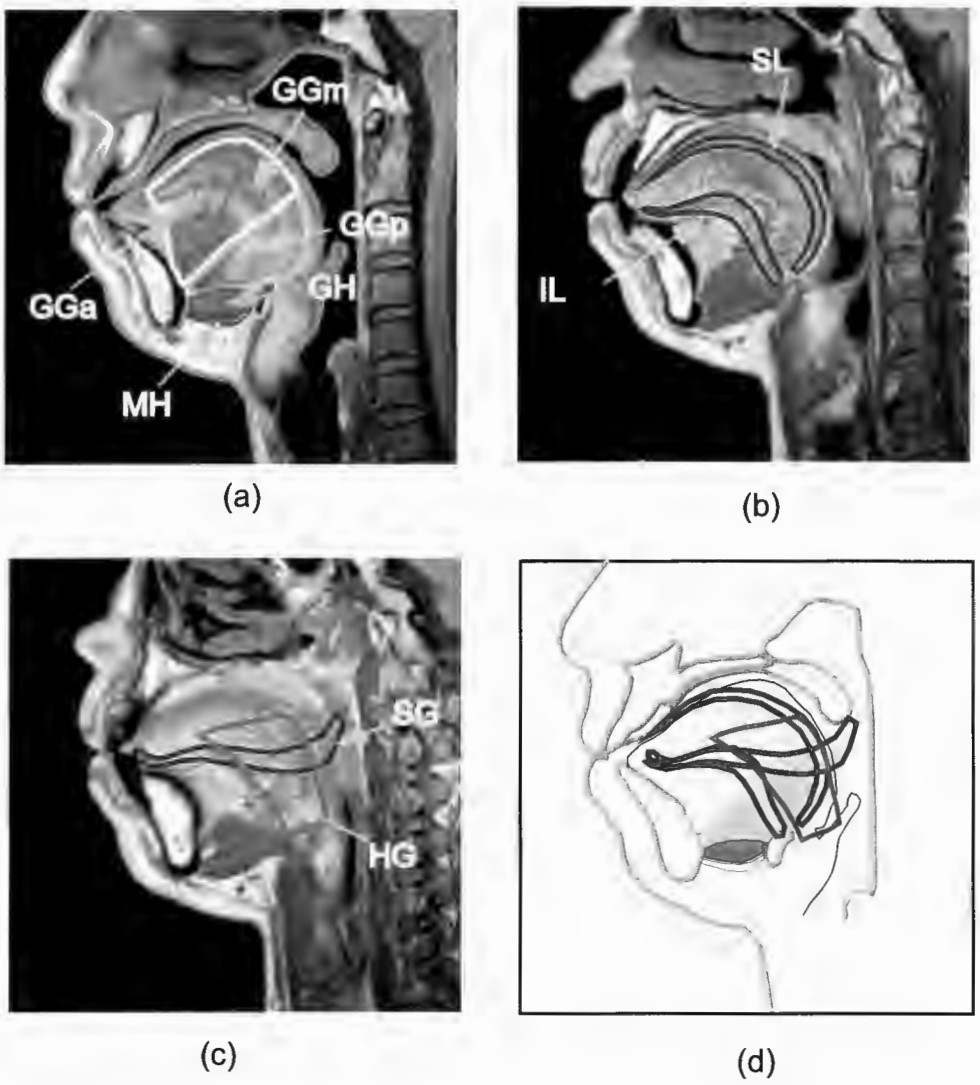


Figure 4 Extraction of tongue muscles and outline of the vocal tract: (a) midsagittal plane, (b) parasagittal plane (0.6 cm), (c) parasagittal plane (1.5 cm), and (d) superimposed view of the extracted outlines.

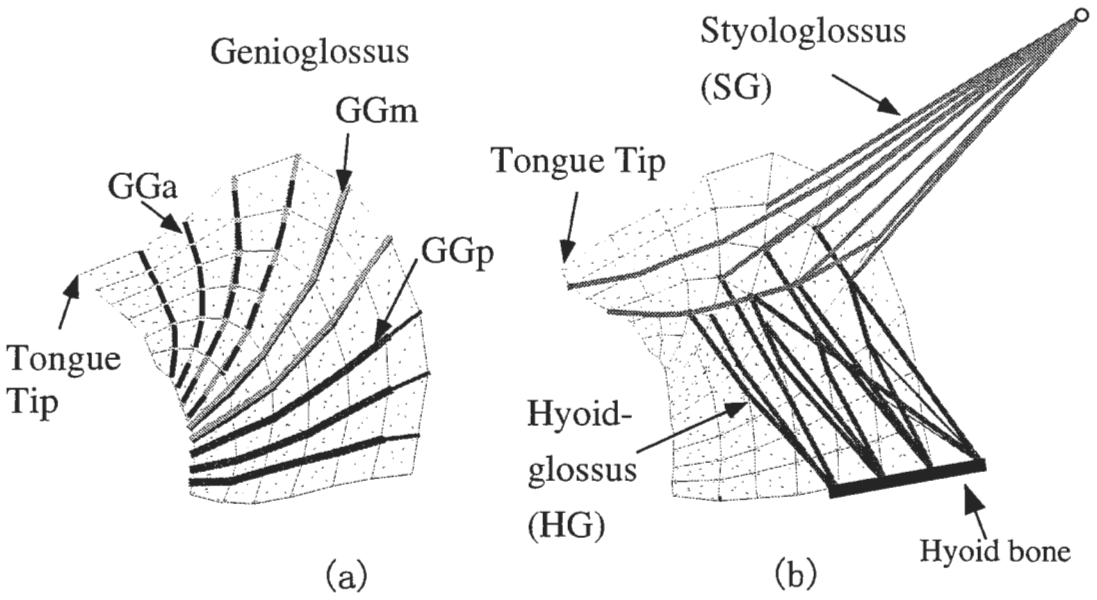


Figure 5 Structure of extrinsic muscles of tongue model: (a) three bundles of genioglossus muscles (GGa, GGm and GGp) on midsagittal plane and (b) hyoglossus (HG) and styloglossus (SG) on parasagittal plane.

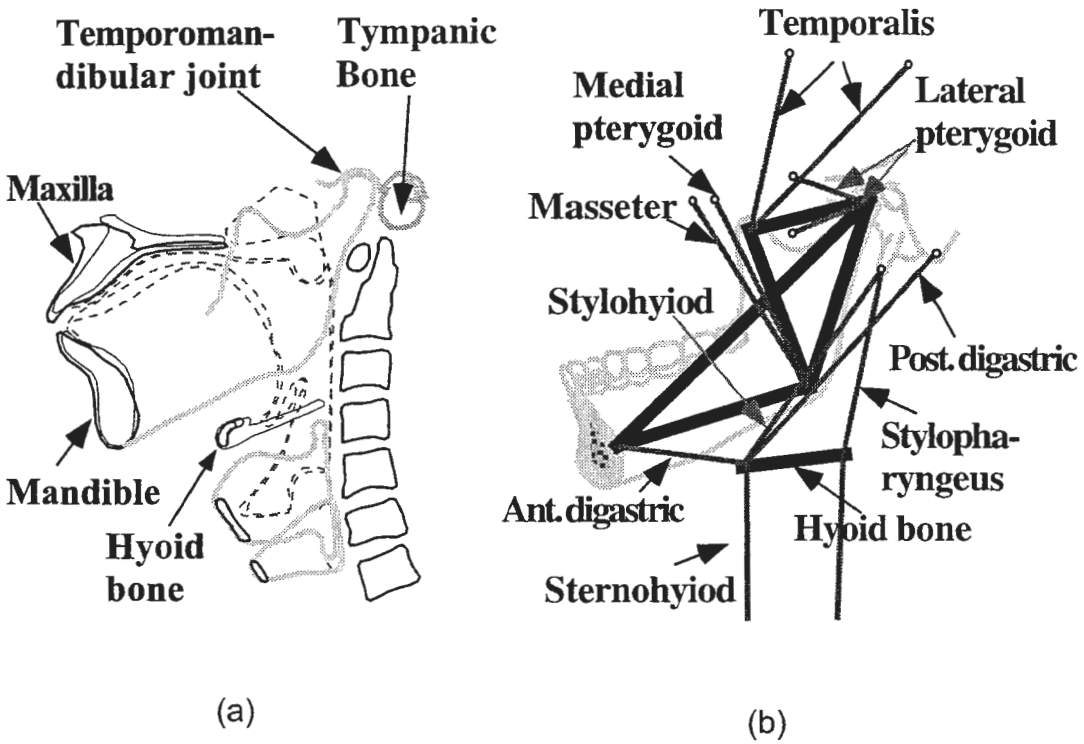


Figure 6 Modeling of the rigid organs based on MR images: (a) extracted framework of bony organs, and (b) model of the mandible and hyoid bone with related muscles (structure of the concerned muscles in reference to anatomical literature).

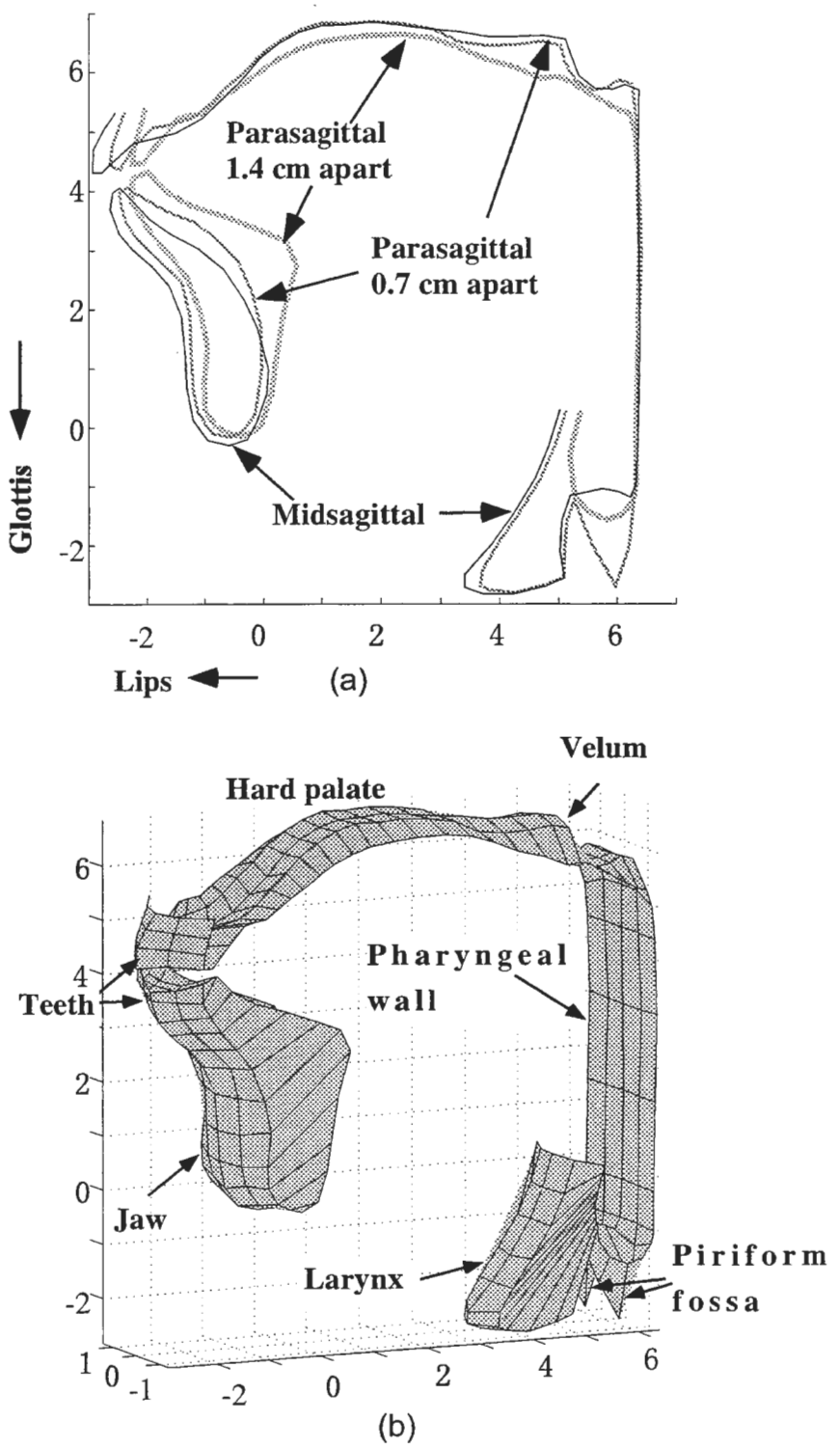


Figure 7 Modeling of the vocal tract wall: (a) extracted outlines of vocal tract wall based on MR images, and (b) reconstructed surface of vocal tract walls assuming symmetric left and right sides (dimensions in cm).

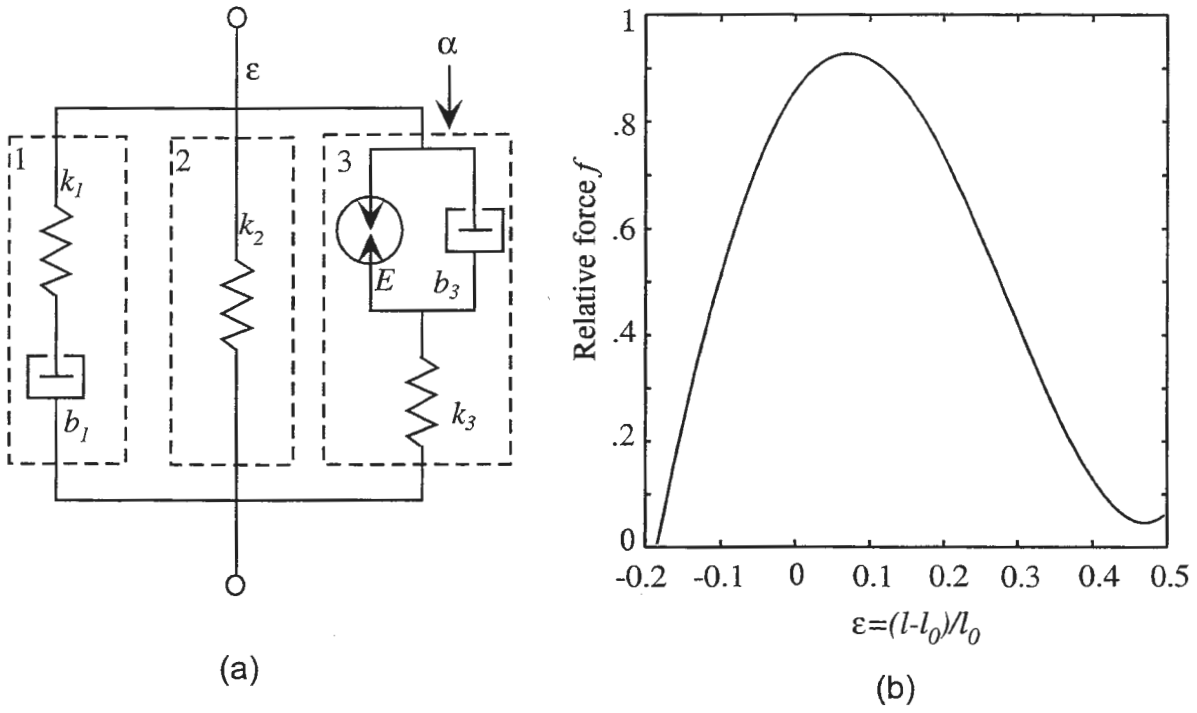


Figure 8 Muscle modeling: (a) a general model of muscle: k and b are stiffness and dashpot, E is contractile element. (b) generalized force varies with stretch ratio ε. α is a coefficient of the active part, depending on muscle thickness and activation level.

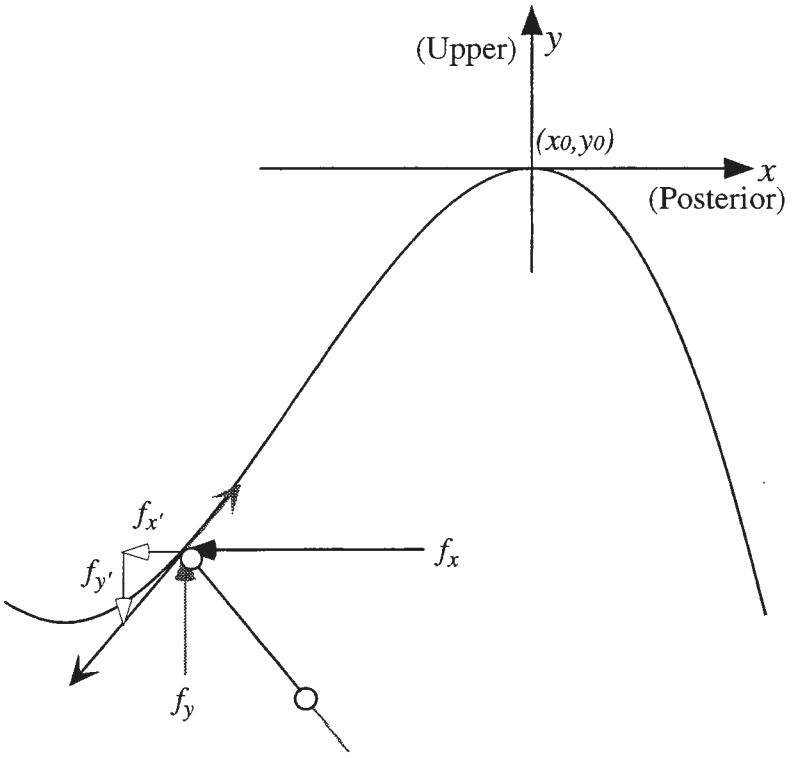


Figure 9 Model of the articular groove of the temporomandibular joint and forces applied on the condyle. Force is redistributed at contact point of the condyle with the path. Dark lines shows forces related to x-direction and gray lines for y-direction. fx and fy are initial inputs. fx' and fy' are redistributed forces.

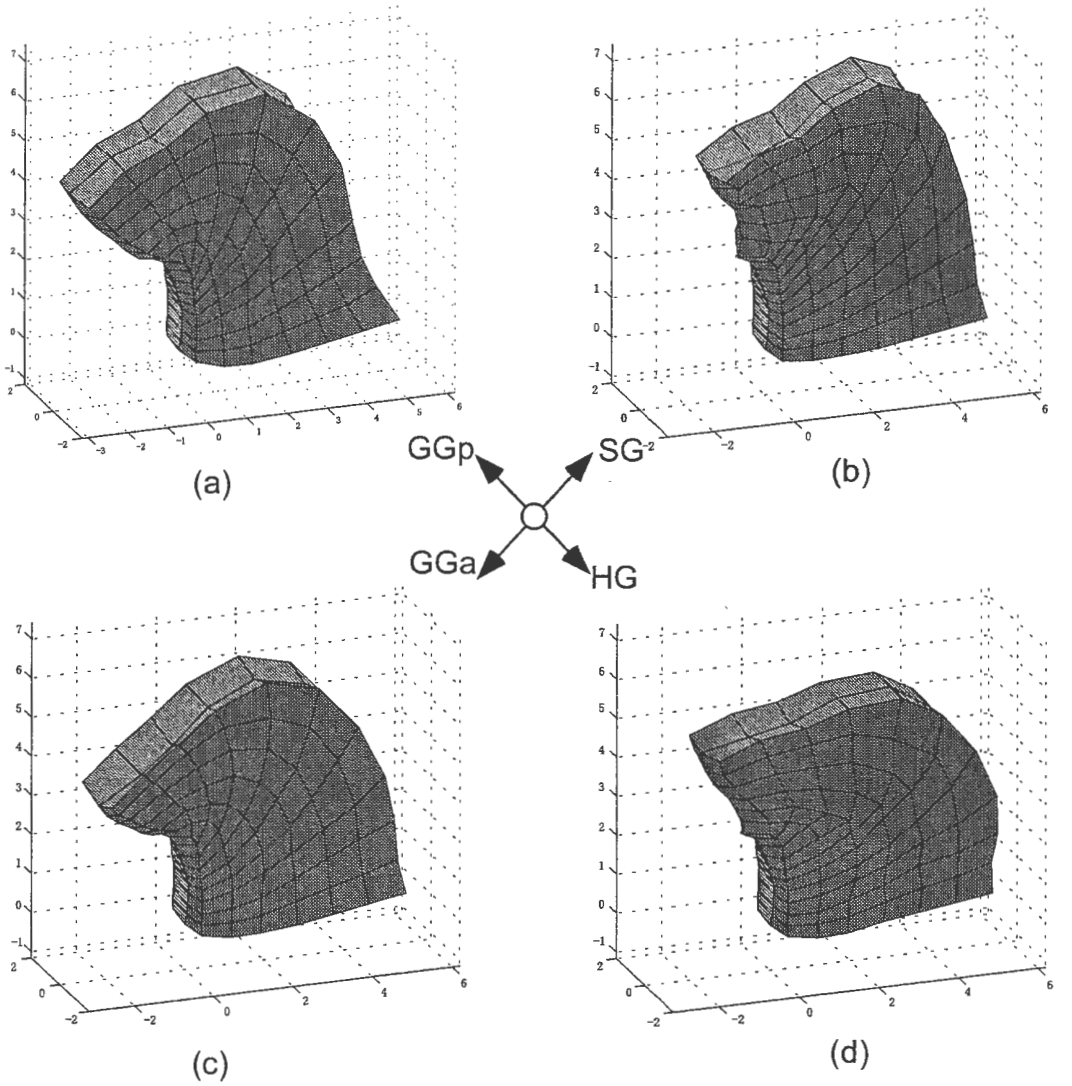


Figure 10 Tongue deformations by contracting extrinsic tongue muscles: (a) tongue dorsum advances and rises by GGp, (b) rises and retracts by SG, (c) lowers and grooves by GGa, and (d) retracts and lowers by HG. Cross arrow indicates the direction of tongue movements by indicated muscles.

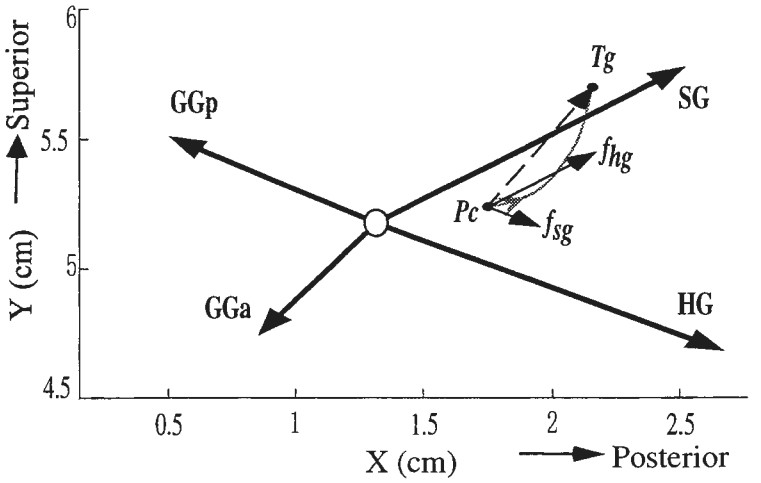
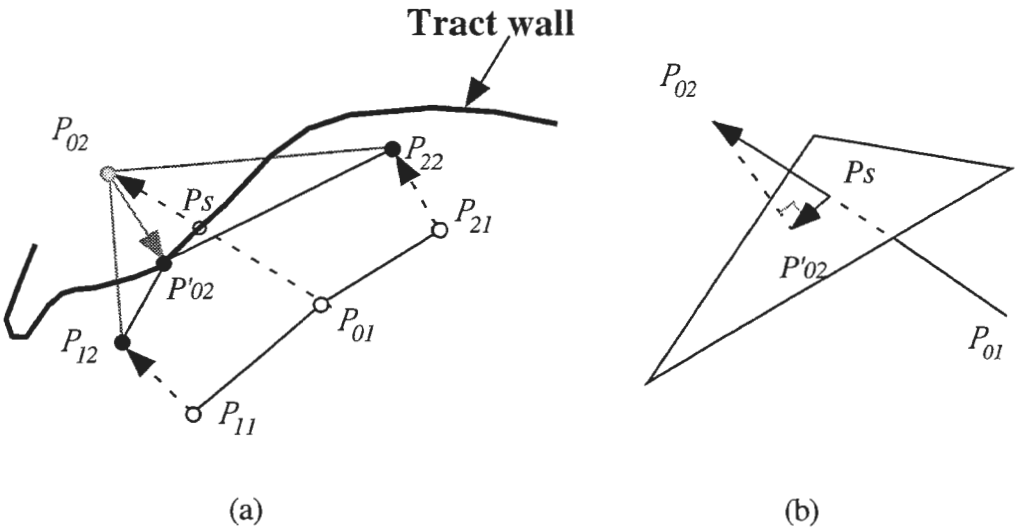


Figure 11 Muscle workspace of the tongue showing a method to compute muscle contraction force. ○: initial position of the tongue; *Pc*: current position of the tongue, *Tg*: articulatory target. *fsg* and *fhg* are positive projections of vector of *Pc* to *Tg* in muscle workspace. Gray line shows trajectory from *Pc* to *Tg*, and gray arrow indicates the direction at current step.



(a) (b)

Figure 12 Modeling of tongue collision on vocal tract wall: (a) deformation during collision and (b) new position of the node on the wall. *Pij:* Position of node *i* at time *j*.

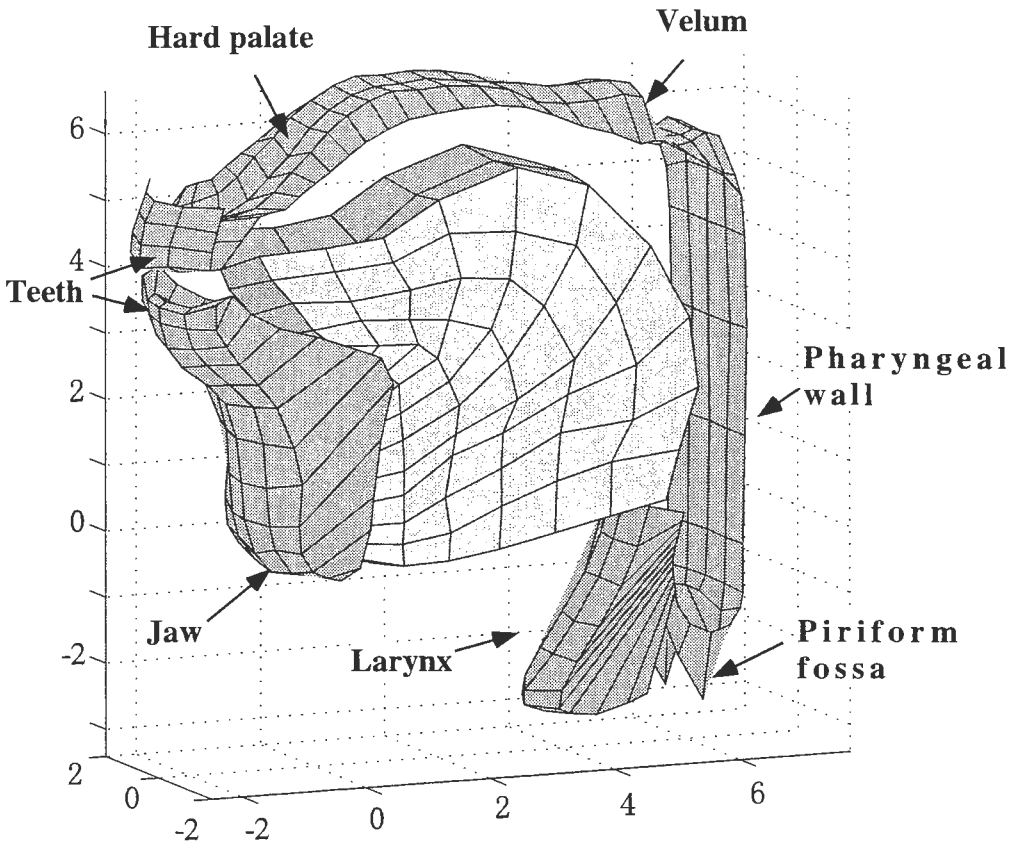


Figure 13 An oblique view of three-dimensional model of the speech organs. All dimensions are in cm.

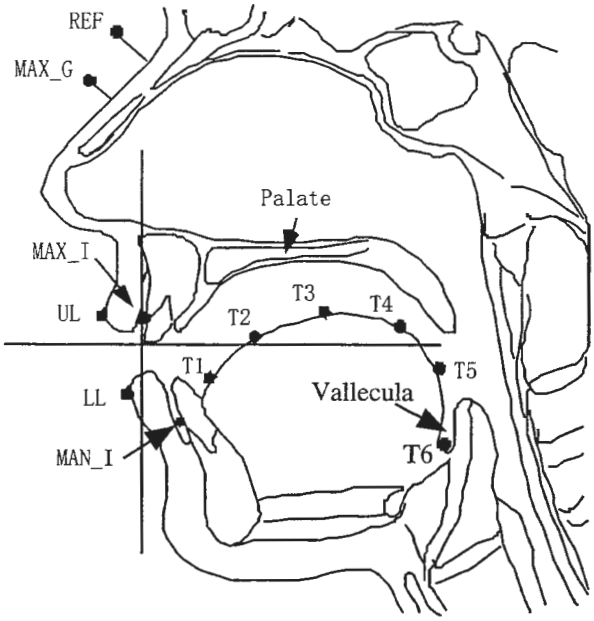


Figure 14 Pellet placements for x-ray microbeam experiments.

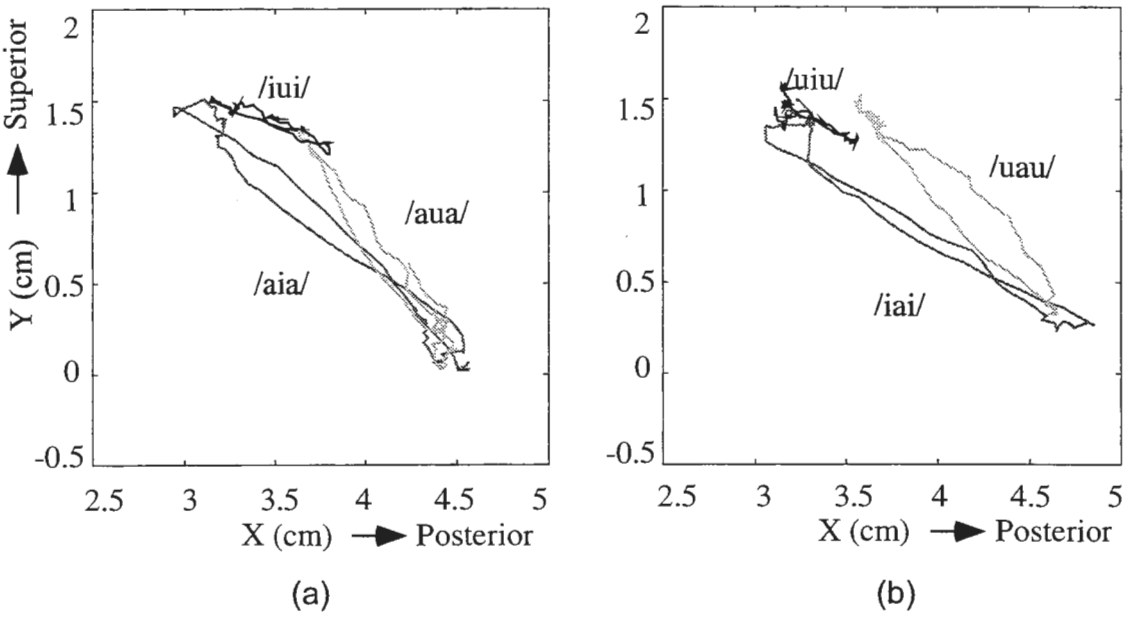


Figure 15 Observed trajectories of the tongue position (average of T1-T3) during three types of vowel sequences.

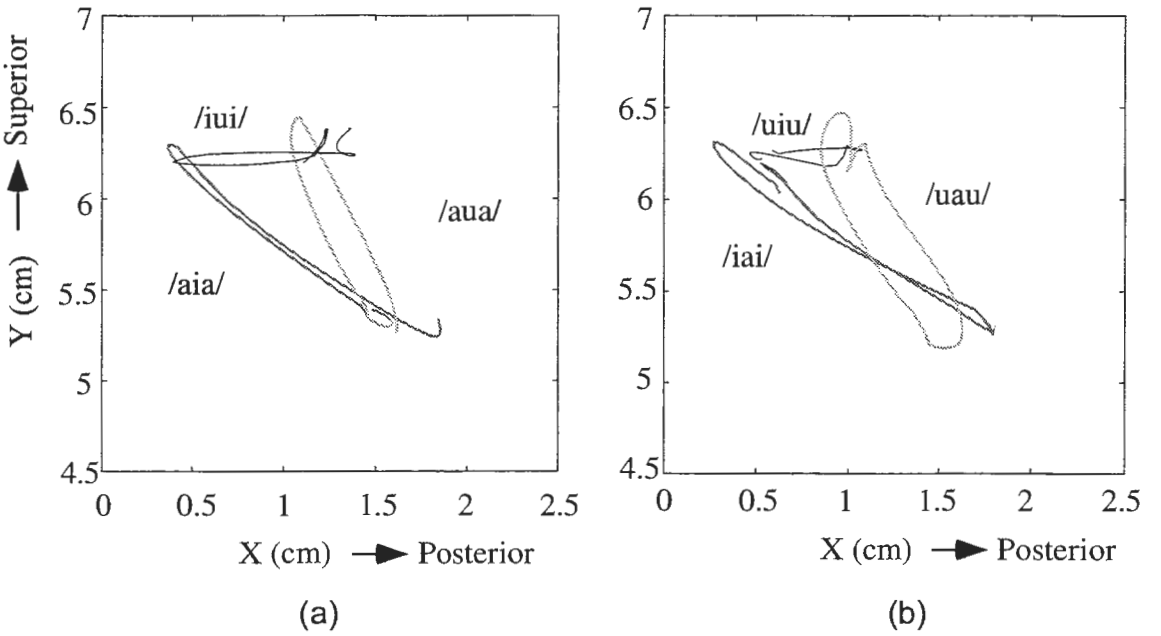


Figure 16 Generated trajectories of the tongue position (average of the same region as that in Fig. 15) for the same vowel sequences.

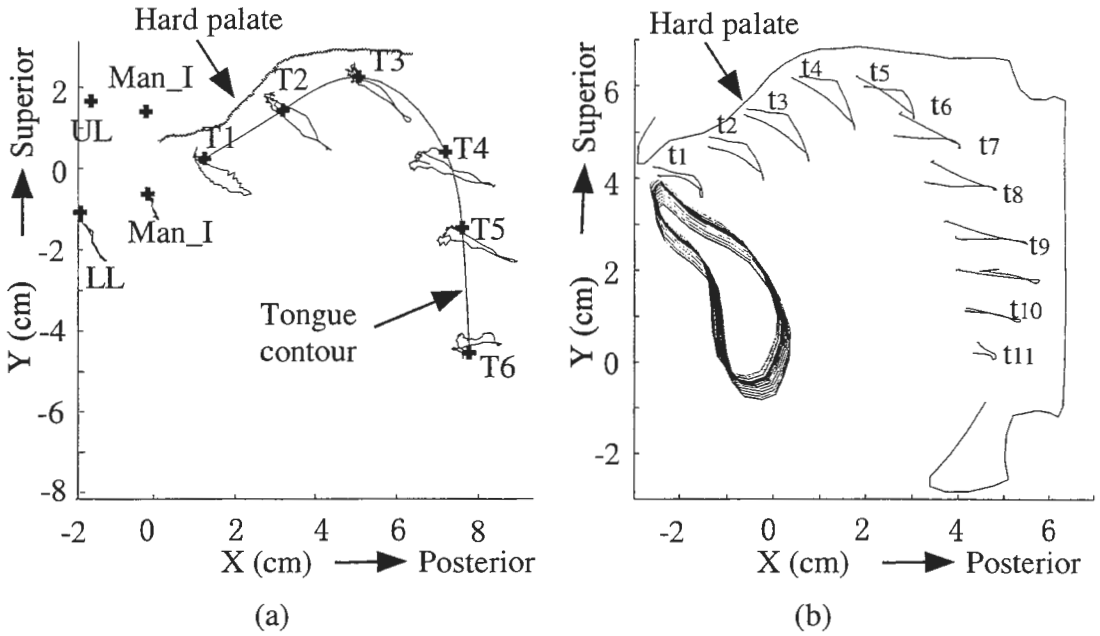


Figure 17 Trajectories of surface points of the speech organs during /iai/ in observations and simulation: (a) movements of pellets (see Fig. 14 for details), and (b) movements of model's surface nodes.

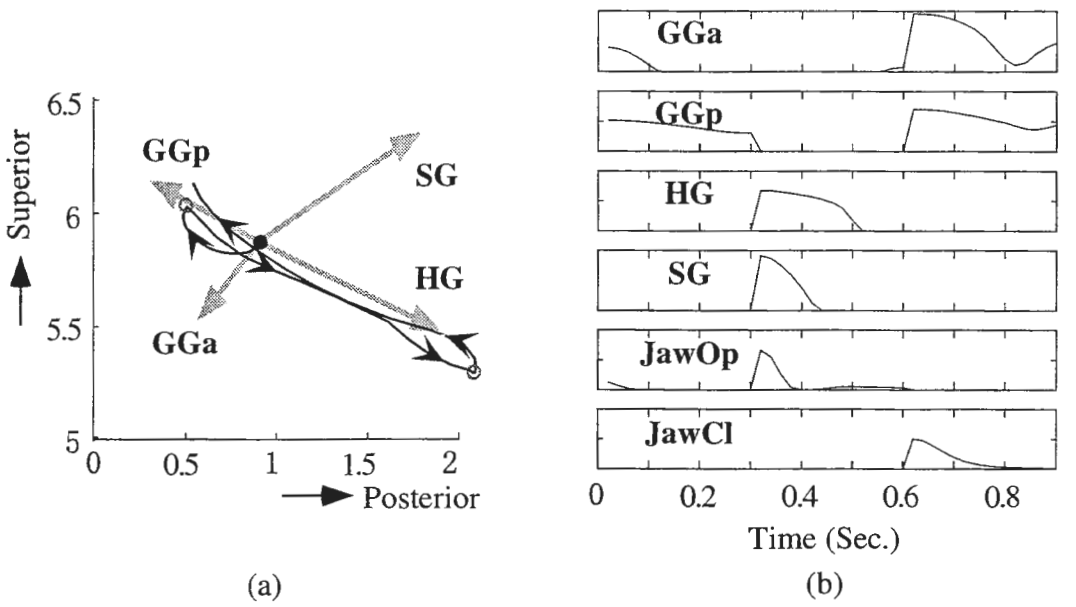


Figure 18 Model control using an articulatory target sequence: (a) articulatory trajectory of tongue position in tongue muscle workspace, and (b) generated muscle activation signals, where the signals of jaw opener and closer groups were obtained similarly from jaw muscle workspace.