A Physiological Articulatory Model for Simulating Speech Production Process

(*A Physiological Model for Speech Production*)

(to be published in Acoustical Science and Technology, Vol. 22, 6)

Jianwu DANG and Kiyoshi HONDA

ATR Human Information Processing Research Labs,

2-2 Hikaridai Seika-Cho, Soraku-gun, Kyoto, Japan, 619-0288

jdang@atr.co.jp and Honda@atr.co.jp

Abstract

A physiological articulatory model has been developed to simulate the dynamic actions of speech organs during speech production. This model represents the midsagittal region of the tongue, jaw, hyoid bone, and the vocal tract wall in three dimensions. The soft tissue of the tongue is outlined in the midsagittal and parasagittal planes of MR images obtained from a male Japanese speaker, and constructed as a 2-cm thick layer. The palatal and pharyngeal walls are constructed as a hard shell of a 3-cm left-to-right width. The jaw and hyoid bone are modelled to yield rotation and translation motions. The muscle structure in the model is identified based on volumetric MR images of the same speaker. A fast simulation method is developed by modeling both the soft tissue and rigid organs using mass-points with two types of links: viscoelastic springs with a proper stiffness for connective tissue, and extremely high stiffness for bony organs. Muscle activation signals are generated by a model control strategy based on the target-reaching task, and then fed to drive the model to approach the targets. The model demonstrated realistic behaviors similar to coarticulation in human speech production (Dang and Honda, 1998, 1999, 2000).

Keyword: Speech production, Articulatory model, Physiological model, Tongue model, Articulatory target control, MRI.

INTRODUCTION

Since the anatomical and biomechanical properties of human speech organs influence articulatory kinematics, speech motor control and

sound patterns of languages, a detailed knowledge of speech production mechanisms is essential for speech communication studies. Because it is not always possible to experimentally uncover all details of the human speech organs, it is necessary to use models to understand speech production characteristics. A number of geometrical or parametrical articulatory models have been developed to simulate the mechanisms of speech production in order to examine the influences of the anatomical and biomechanical properties on speech sounds. The following is a brief summary of modelling studies.

One of the first physiological models of the human tongue was constructed by Perkell [1]. It was a two-dimensional (2D) projection of the tongue in the sagittal plane composed of a lumped parameter and a lumped force system, equivalent to the finite element method (FEM). Using the FEM approach, 3D tongue models were investigated by Hashimoto and Suga [10], Kakita et al. [15], and Kiritani et al. [16]. In these models, neither the inertial component nor the effects of geometric nonlinearities were represented. To account for the factors, Wilhelms-Tricarico proposed a rigorous method for modelling the soft tissue and then built a 3D tongue model [28]. A 2D biomechanical tongue model was built by Payan and Perrier using the FEM [22]. Their model is used to produce V-V sequences according to one of the common motor control theories, the Equilibrium Point Hypothesis (EPH).

Hirai et al. developed a 2D physiological model unifying the soft tissue and rigid body of the tongue, jaw and laryngeal structures [12, 13]. The soft tissue of the tongue was modelled by FEM based on MRI data from a male speaker. The rigid organs (jaw, hyoid bone, thyroid cartilage,

and cricoid cartilage) were connected by muscles to form a mass-spring network system having contractile units. The dynamic balance of forces and moments was used as a mechanical principle to interface the soft and rigid structures. This model successfully reproduced biomechanical interaction between the tongue and larynx, corresponding to the observations of the MRI data.

On the other hand, Sanguineti et al. employed 2D model of the tongue, jaw, hyoid bone, and larynx to develop a control strategy based on the EPH (λ model) [25]. The dynamic behaviour of the whole system was specified by its global kinetic energy and potential energy function. Their results showed that all movements could be approximated as linear combinations of elementary motions. They noted that the soft tissue and rigid organs have quite different dynamic behaviors and the dynamic effects that occurred at the interface between the soft tissue and rigid organs were not negligible in modelling speech-like movements.

Generally speaking, most of the studies of physiological articulatory models, especially 3D models, focus on theory and methodologies for constructing a model. The aim of this study is to develop a physiological articulatory model that can be used as a practical tool in speech research. We start from careful construction of a subject-specific 3D model that can demonstrate human speech articulation with a time-efficient algorithm and a practical target-based control strategy. In this paper, we focus on construction of a partially 3D physiological articulatory model and development of a target-based control strategy.

I. DESIGN OF THE ARTICULATORS

To construct a subject-specific model, the shape of the tongue and

the contour of the rigid organs are extracted based on volumetric MRI data, which obtained from a male Japanese speaker using a standard spin echo method. A set of high-resolution volumetric MR images with a smaller slice thickness is used to identify muscle structures.

A. Modeling of the Tongue Body

During natural speech the tongue either forms lateral airways by narrowing the tongue body, or, makes a midsagittal conduit by contacting the palate with lateral parts and/or grooving tongue. This is seen with some consonants and high-front vowels. The model of the tongue is designed to perform basic 3D deformations such as midsagittal grooving and lateral airway formation. A trade-off between computational cost and model similarity resulted in a partial 3D model, with a 2-cm thick sagittal layer rather than a full 3D model.

1. Derivation of the governing equations

The soft tissue of the tongue has commonly been modeled using the finite element method (FEM) [15, 28]. In this modeling, we started to derive the governing equations of the tongue tissue based on FEM, and then realized the governing equations using a mass-spring network with an appropriate simplification.

In the finite element analysis, the soft tissue of the tongue body is approximated as an assemblage of discrete finite elements interconnected at nodal points on the element boundaries. The displacement u measured in a local coordinate system x, y, z within each element is assumed to be a function of the displacement at all the nodal points on the element. Therefore, for element m we have

$$u^{(m)}(x,y,z) = H^{(m)}(x,y,z)X^{(m)} \tag{1}$$

where $H^{(m)}$ is the displacement interpolation matrix, the superscript *m* denotes element *m,* and $X^{(m)}$ is a vector of global displacement components in three dimensions for all nodal points. The corresponding element strain is given by

$$\varepsilon^{(m)}(x,y,z) = D^{(m)}(x,y,z)X^{(m)} \tag{2}$$

where $D^{(m)}$ is the stain-displacement matrix; the rows of $D^{(m)}$ are obtained by appropriately differentiating and combining the rows of $H^{(m)}$. According to the Hamilton's principle, the stationary equilibrium is reached when the derivative of the total energy, consisting of the strain energy, the kinetic energy, and the work done by external forces, is equal to zero. With a common assumption that the tongue tissue can be approximated as an isotropic material, the derived equations of equilibrium governing the linear dynamic response of a finite element system are

$$M\ddot{X} + B\dot{X} + KX = F \tag{3}$$

where *M, B,* and *K* are the mass, damping, and stiffness matrices; *F* is the vector of externally applied loads; *X*, $\dot{X}$, and $\ddot{X}$ are the displacement, velocity, and acceleration vectors of the finite element assemblage.

$$\begin{aligned} M &= \sum_m \int_{V^{(m)}} \rho H^{(m)T} H^{(m)} dV^{(m)} \\ B &= \sum_m \int_{V^{(m)}} b H^{(m)T} H^{(m)} dV^{(m)} \\ K &= \sum_m \int_{V^{(m)}} D^{(m)T} C D^{(m)} dV^{(m)} \\ F &= \sum_m \left(\int_{V^{(m)}} H^{(m)} f^B dV^{(m)} + \int_{S^{(m)}} H^{S(m)} f^S dS^{(m)} \right) + f^L \end{aligned} \tag{4}$$

where ρ is the mass density, and b is the damping property parameter; C is the generalized stress-strain matrix, which depends only on the Young's modulus and Poisson's ratio for an isotropic material. f^B, f^S, and f^L are the body force, surface force and concentrated force, respectively.

The dumping matrix B and the stiffness matrix K are sparse matrices. From the displacement interpolation and the strain-displacement matrices, it is easy to find that element e_{ij} of the matrices is non-zero only if nodal point i is adjacent to nodal point j. To materialize the matrices in a brief formation, viscous and stiffness components of the non-zero elements are represented by using a viscoelastic spring to connect the adjacent point pairs. The properties of the matrices B and K can be correctly represented if appropriate values are selected for the viscoelastic springs. According to a lumped mass matrix, the masses are distributed in the nodal points. Thus, the soft tissue of the tongue body is modeled as a mass-spring network.

According to literature [9], there are three types of models for representing a viscoelastic material: the Voigt model, Maxwell model, and Kelvin model. The Voigt model consists of a spring parallel to a dashpot, while the Maxwell model consists of spring cascaded as a dashpot. The Kelvin model is a combination of the first two models. The relation between force F, displacement u and velocity $\dot{u}$ is described in 5(a) for the Voigt model, and 5(b) for the Maxwell model.

$$\begin{aligned} F &= ku + b\dot{u} \qquad (a) \\ u &= \dot{F}/k + F/b \qquad (b) \end{aligned} \tag{5}$$

where k and b denote stiffness and viscous coefficients, respectively. Comparing (5) with (3), it is obvious that the Voigt model is more convenient to be incorporated in the motion equation. When a force is applied to the Voigt model, a deformation gradually builds up as the spring shares the load. After the force is removed, the dashpot displacement relaxes exponentially, and the original length is restored from the deformation.

2. Extraction of tongue shapes based on MRI data

The MRI data used to replicate the tongue and other speech organs consist of 15 sagittal slices. A 30cm×30cm field of view was digitally represented by a 256×256 pixel matrix for each slice. The relaxation time (TR) was 500 ms and the excitation time (TE) was 20 ms. The slices were 0.7 cm thick with no gap or overlap. The whole data set was processed using commercial software (VoxelView) on a workstation, IRIS Indigo, into a 3D volumetric image with a 0.1cm ×0.1cm ×0.1cm voxel. Tongue tracings were derived from the reconstructed 3D data. The tongue shape of a Japanese vowel [e] is chosen as the initial shape of the model. The outlines of the tongue body are extracted from two sagittal slices; one is the midsagittal plane and the other is a plane 1.0cm apart from the midsagittal on the left side. Assuming that the left and right sides of the tongue are symmetrical, the outline of the right side is a copy of that of the left side. Figure 1 shows the extracted outline of the tongue and the corresponding MR images. The contours of the tongue show some differences in the midsagittal and parasagittal planes. The causes of the differences are the tongue groove in the anterior portion, and the vallecula, the valley of the epiglottis, in the posterior portion. The

Figure 1

epiglottis was not included in the tongue model, but its volume is considered in calculating area functions of the vocal tract for synthesis of speech sounds. The outline of the tongue root in the parasagittal plane is slightly exaggerated for considering the attachment of the tongue musculature to the greater horns of the hyoid bone.

As mentioned in the preceding section, we use a mass-spring network as a basis to approximate the soft tissue of the tongue. The basic structure of the network is adopted from the fiber orientation of the genioglossus muscle. The midsagittal region of the tongue that includes this muscle is represented by three sagittal planes. The tongue tissue in each plane is divided into ten radial sections that fan out from the genioglossus' attachment on the jaw to the tongue surface. In the perpendicular direction, the tongue tissue is divided into six sections concentrically. The mesh pattern obtained from this segmentation is shown in Fig. 1. In this model, the mass-points are located at the junctions of the mesh lines, and viscoelastic springs connect each point to the surrounding points, in which one point can connect with up to 26 points.

The current model of the tongue tissue consists of 120 segmented units (meshes). Each mesh is defined by 12 edges with eight vertices. Supposing that the mass of a mesh is equally distributed on the eight vertices, the total mass of a node is the summation of the masses of the vertices that share this node. The mass per unit volume is 1.0 g/cm^3 for the tongue tissue, which is the same as that of water [25]. As a result, the total mass of the 2-cm thick sagittal layer of the tongue was about 56 g (the volume calculation is described in section II C).

The mechanical parameters for the spring and the dashpot reported in the previous studies have differed widely. The stiffness ranged from 10^4-10^6 dyne/cm^2, and the viscosity from 10^5-10^7 dyne•s/cm^2 [24]. In the present model, both parameters were chosen in the same order: 1.98×10^5 dyne/cm^2 for the stiffness and 2.25×10^5 dyne•s/cm^2 for the viscosity. Since there are nine connections in a mesh of the tongue tissue, the parameter values for a viscoelastic spring are one-ninth of these values. Table 1 shows the parameters of the mass, viscosity and stiffness used in this model.

Table 1

B. Arrangement of the Tongue Muscles

To realize the subject-specific customization, we examined the anatomical arrangement of the major tongue muscles based on a set of high-resolution MR images obtained from our target speaker. The data set consists of 40 sagittal slices with a 0.35-cm thickness and a 0.05-cm overlap acquired during a rest position. A 25cm×25cm field of view for each slice was digitally represented by a 256×256 pixel matrix. The excitation time (TE) was 15 ms and the relaxation time (TR) was 620 ms. The boundaries of muscles were traced in each slice, and then superimposed together so that the major muscles could be identified.

Figure 2

Figure 2 shows an example of the muscle outlines projected in three sagittal planes, in which the specific muscles were easily distinguished. Figure 2(a) is the midsagittal plane showing the genioglossus (GG) and the geniohyoid (GH). Figure 2(b), being 0.6 cm apart from the midsagittal, depicts the superior longitudinalis (SL), and inferior longitudinalis (IL). Figure 2(c), which is 1.5 cm apart from the midsagittal, indicates the hyoglossus (HG) and the styloglossus (SG).

Figure 2(d) combines the muscle tracings. The other intrinsic muscles (transversus and verticalis) were not identifiable in the MR images. The orientation of the tongue muscles was also examined with reference to the literature [19, 26, 27].

Since the muscles have a certain thickness in the left-right direction, it is difficult to model them within one plane alone. The largest extrinsic muscle GG shown in Fig. 2(a), for example, runs midsagittally in the central part of the tongue. To realize its structure realistically, the muscle is not only arranged in the midsagittal plane, but also arranged in the parasagittal planes. The muscles traced in Fig. 2 (b) are designed in both the midsagittal and parasagittal planes. The muscles in Fig. 2 (c) are designed in the parasagittal planes alone. Figure 3 shows an example of the tongue muscle arrangement used in the proposed model. Figure 3(a) shows the arrangement for the GG. Since the GG is a triangular muscle and different parts of the muscle exert different effects on tongue deformation, it can be functionally separated into three segments: the anterior portion (GGa) in the dashed lines, the middle portion (GGm) in the dark lines, and the posterior portion (GGp) in the gray lines. The line thickness represents the approximate size of the muscle units, and the thicker the line, the larger the maximal force generated. Figure 3(b) and 3(c) show the arrangement of other extrinsic muscles, the hyoglossus (HG) and styloglossus (SG), in the parasagittal plane, where the thickest line represents the hyoid bone. In addition, two tongue-floor muscles, geniohyoid and mylohyoid, are also shown in the parasagittal planes. Note that the geniohyoid was also modeled in the midsagittal plane though it is not plotted in the figure. The top points of the bundles of the mylohyoid muscle are attached to the medial surface of the mandibular

Figure 3

body. All the muscles are designed to be symmetrical in the left and right sides. The lower panels show four intrinsic muscles. The transversus muscle runs in the left-right dimension, and its location is plotted in the midsagittal plane with the star markers. Altogether, eleven muscles are included in the tongue model (Table 2).

C. Modeling of the Rigid Organs

The outlines of the rigid organs (the jaw and hyoid bone in the present work) were also traced from the MRI data for the target subject. Although the bony organs were not visualized in MR images due to the lack of water, the contour of the organs can be identified in MR images when soft tissues surround them. The data used to extract the contour of the rigid organs was the same set as that described in Section I A. Figure 4 (a) shows the bony framework extracted from the midsagittal plane and a parasagittal plane with a 0.7-cm interval. The gray thick lines show the contours of the organs drawn with reference to the anatomical literature, and the dashed lines are the extracted boundary of the soft tissue. The thick dark lines show the rigid organs traced on the midsagittal plane, and the thin lines for the organs on the parasagittal plane.

Figure 4

Figure 4 (b) shows the model of the jaw and the hyoid bone. The right half of the mandible is drawn in the background using pale gray lines. The model of the jaw 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 model of the jaw is combined with the tongue model at the mandibular symphysis. The model of the hyoid

bone has three segments corresponding to the body and bilateral greater horns. Each segment is modelled by two mass-points connected with a rigid beam for each side. Yamazaki investigated the weight of the cranium and mandible using 92 dry skulls from Japanese specimens. His result showed that the weight of the male jaws was around 90g [29]. According to this literature, the equivalent mass of the living jaw is roughly estimated to be 150g included water and surrounding tissue. To evaluate the mass for the hyoid bone, the structure of the hyoid bone was extracted and measured using volumetric computer topographic data. The volume of the hyoid bone was about 2.5 cm^3 for a male subject. Based on this measurement, an equivalent mass was set at 5 g for the hyoid bone. Note that both of the masses are much smaller than those used in [25]. The masses are equally distributed in the body nodes. In the present model, rigid beams are also treated as viscoelastic links so that they can be integrated with the soft tissue in the motion equation. Their values are about ten thousand times greater than those used for the soft tissue.

The eight muscles with thin lines in Fig. 4(b) are incorporated in the model of the jaw-hyoid bone complex, where the structure of the muscles was based on the anatomical literature [27]. The small circles indicate the fixed attachment points of the muscles. Since the other rigid organs below the hyoid bone, such as the thyroid and cricoid cartilages, are not included in the present model, two viscoelastic springs are used as the strap muscles. The temporalis and lateral pterygoid are modeled as two units to represent their fan-like fiber orientation. The anterior and posterior bellies of the digastric muscle are modeled to connect the hyoid bone at a fixed point. All these muscles are modeled symmetrically on

the left and right sides. Jaw movements in the sagittal plane involve a combination of rotation (change in orientation) and translation (change in position). There is no one-to-one mapping between muscle actions and kinematical degrees of freedom. However, the muscles involved in the jaw movements during speech can be roughly separated into two groups: the jaw-closer group and the jaw-opener group. Table 2 lists these muscles, where the superscript "o" denotes the muscles belonging to the opener group and "c" for the closer group. The muscles without the superscripts are not activated, but they play an elastic recoil function. Note that the anterior portion and posterior portion of the digastric considered as one muscle in this study.

D. Construction of the Vocal Tract Wall

To form a vocal tract shape, it is necessary to incorporate the vocal tract wall in the model. For this purpose, outlines of the vocal tract wall were extracted from the MRI data described in Sec. I A. Figure 5 shows the extracted outlines and the reconstructed 3D shell for the vocal tract wall and the mandibular symphysis. In Fig. 5 (a), the thin dark lines show the contour of the walls in the midsagittal plane. The pale thick lines indicate the walls in the parasagittal plane on the left side 1.4 cm from the midsagittal plane, and the medium lines for the plane 0.7 cm from the midsagittal plane. With an assumption that the left and right sides are symmetric, we reconstruct a 3D shell of the vocal tract wall and the mandibular symphysis using the outlines with 0.7 cm intervals in the left-right direction. Because of the geometrical complexities, it is impossible to derive an analytic function for the shell surfaces. For this reason, the shell surfaces of the vocal tract wall and the mandibular

Figure 5

symphysis are approximated using a number of triangular planes. Figure 5(b) shows the reconstructed vocal tract shell with the tongue body in an oblique-front view. The shell surface of the tract wall is compounded using 432 triangular planes, and 192 triangular planes are used for the mandibular symphysis, where triangles are replaced by quadrilaterals in the figure for a concise description. The piriform fossa, which has bilateral cavities and behaves as side branches in the vocal tract [3], is also combined in this shell model. This figure demonstrates a whole image of the proposed model, which consists of the tongue, jaw, hyoid bone, and vocal tract wall.

II. COMPUTATIONAL METHODS

As described above, both the soft tissue and rigid organs are integrated into a motion equation system using a mass-spring network. This section describes the computational methods used in the compound model, and the constraints for the volume of the soft tissue and for the movement of the rigid organs.

A. Solution of the Motion Equations

The motion equation of the mass-spring network is described as a second order differential equation of (3). To obtain a high stability in solving this differential equation, an implicit approach, the Houbolt integration method, is employed in the finite difference expansions [2]. Thus, equation (3) at time $t+h$ can be rewritten as

$$M\ddot{X}(t+h) + B\dot{X}(t+h) + KX(t+h) = F(t+h) \tag{6}$$

Using a backward-difference method, we obtain the solution of $X(t+h)$,

$$\begin{aligned}(2M + 11Bh/6 + Kh^2)X(t+h) = {} & h^2F(t+h) + (5M + 3Bh)X(t) \\ & - (4M + 3Bh/2)X(t-h) + (M + Bh/3)X(t-2h)\end{aligned} \tag{7}$$

M is a diagonal matrix consisting of the masses of all the mass-points within the model (the tongue, jaw, and hyoid bone). B and K are the damping matrix and stiffness matrix, respectively, whose elements $k_{i,j}$ and $b_{i,j}$ are represented by the following expressions.

$$
\begin{aligned}
&k_{3i+u,3i+v} = \sum_{c=1}^{cn} kk_{ic} r_{ic}(u) r_{ic}(v), \qquad k_{3c+u,3i+v} = kk_{ic} r_{ic}(u) r_{ic}(v), \\
&b_{3i+u,3i+v} = \sum_{c=1}^{cn} bb_{ic} r_{ic}(u) r_{ic}(v), \qquad b_{3c+u,3i+v} = bb_{ic} r_{ic}(u) r_{ic}(v), \\
&u = 0,1,2; \quad v = 0,1,2.
\end{aligned} \tag{8}
$$

where bb_{ic} *and* kk_{ic} are viscous and stiffness components of the spring connecting node i and c, and $r_{ic}(u)$ denotes the direction cosine from node i to node c: u equals 0 for x-direction, 1 for y-direction, and 2 for z-direction. Since the direction cosine $r_{ic}(u)$ varies with time, the viscosity and stiffness matrices are time varying. $F(t)$ in (6) denotes the external forces on the nodal points.

B. Computation of External Forces

This model involves two kinds of external forces: one is generated by muscle contraction and the other is induced by soft tissue contact with the rigid boundaries. The former is the source force to drive the model. Since the initial shape is assumed to be an unloaded configuration, the gravity is not taken into this computation.

1. Generation of muscle forces

In formulating the generalized model of the muscles, this study adopts a commonly accepted assumption: a force that depends on muscle length is the sum of the passive component (independent of muscle activation) and the active component (dependent on muscle activation). Figure 6 (a) shows a diagram of the rheological model for a muscle unit

Figure 6

proposed by Morecki [20]. The muscle unit consists of three parts that describe the nonlinear property, the dynamic (force-velocity) property, and the force-length property.

Part 1 is a nonlinear spring k_1, which is involved in generating force only when the current length of the muscle unit is longer than its original length. The value of k_1 is selected as k_1=0.05k_0, where k_0 is the stiffness of the tongue tissue. Part 2 consists of the Maxwell body, and is always involved in the force generation. The force generated by this part is determined by two factors: the derivative of the muscle length and the previous force of this branch. The value of k_2 is set to be twice that of the tongue tissue, while b_2 is on the order of one tenth that used in the tongue body. Part 3 of the muscle unit corresponds to the active component of the muscle force, which is Hill's model consisting of a contractile element parallel to a dashpot, cascaded with a spring [11]. This part generates force when a muscle is activated. In model computations, we use a force-length function of the muscle tissue. The force-length function was derived to match the simulation and empirical data by using the least square method [20]. The function arrived at a fourth-order polynomial of the stretch ratio of the muscles,

$$\sigma_3 = 22.5\varepsilon^4 + 3.498\varepsilon^3 - 14.718\varepsilon^2 + 1.98\varepsilon + 0.858 . \qquad (9)$$

where σ_3 is the stress of branch 3 of the muscle unit, and ε is the stretch ratio of the length increment of the muscle to its original length. This empirical formula is valid for $-0.185 < \varepsilon < 0.49$, where the active force is assumed to be zero if ε is out of the range. As shown in Fig. 6(a), a coefficient of α is used as a gain for the active part to generate the maximum force. In this model, α is chosen to be 6000 for all

muscular units, which is determined from model simulations. Figure 6(b) shows the relationship between the stretch ratio of the muscular unit and the generated force including the passive force. This figure demonstrates the force-length characteristic of the muscle model.

2. Force redistribution

The jaw in this model performs translation and rotation motions in the sagittal plane. Since the mandibular condyle slides along the articular groove, force redistribution takes place in the tempromandibular joint during jaw movement. When the condyle receives a force, the force is decomposed into two components. One is parallel to the tangential line (*i.e.*, the slope) of the articular groove at the given point, and the other is consistent with the normal line. The tangential component is responsible for the movements of the jaw while the normal one is counteracted by the surface reaction force. Figure 7 shows force redistribution at the contact point of the condyle with the curved groove. When forces f_x and f_y act on the condyle in the X and Y directions, respectively, each of them decomposes into two force components, one parallel to the slope of the curved groove and one perpendicular to the slope. The dark lines show the components derived from f_x and gray lines indicate the components from f_y. The line with circles shows the summation of the normal forces, which does not contribute to the jaw movements. The lines with a V-shape arrow show the tangential forces. The resultant force is the summation of the tangential forces. The resultant force is further decomposed into $f_{x'}$ and $f_{y'}$ shown in the white arrows, and is then implemented in the motion equations.

Figure 7

C. Constraints of the Motion Equations

To describe the properties of the soft tissue and rigid organs correctly using a mass-spring network, we introduce two constraints in this model: tongue volume constraint and jaw movement constraint.

The tongue body is commonly considered to consist of incompressible tissue. Since a model with mass-spring connections alone lacks incompressible properties, it must have a constraint to maintain the volume of the tongue tissue when the tongue deforms. To do this, it is necessary to calculate the volume for the tongue. Since four adjacent vertices in a mesh are usually not coplanar, there is no analytic expression available. However, there are two (and only two) distinct ways to divide such an eight-cornered mesh into five tetrahedrons and then to obtain its volume [30]. We use the averaged volume of the two types of subdivision in this study to increase the accuracy of the volume during tongue deformation. The volume constraint is to reduce the changes between the current volume $V_j(t)$ and the original volume $V_j(0)$ for each mesh j. This study employs the volume constraint on each mesh instead of on the whole body. This is because the former can avoid the potential risk that the resultant error may be concentrated on one of the meshes.

Jaw movement is not a pure joint rotation because the condyle translates forward as the jaw opens wide [8, 21]. The translation takes place when the condyle slides along the articular groove. To simulate the condyle motion, we assume that the movement of the anatomical center of rotation of the condyle follows a curved path corresponding to the concave articular groove of the tempromandibular joint [17]. The

curved path is approximated using a third-order polynomial

$$y = y_0 - 2.5(x - x_0)^3 - 4(x - x_0)^2, \tag{10}$$

where x and y are the horizontal and vertical coordinates of the center of rotation 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 [21]. Figure 7 shows the curved path of the tempromandibular joint, given in (10). The constraint of the curved path for the condyle is given by minimizing the difference between the actual position, y_i, of the condyle and the position, y_{pi}, predicted from Eq. (10), where i is the index of the two sides.

To combine the above constraints into the main system, we define a cost function for the whole model as in (11):

$$M_c = \| AX - C \|^2 + \gamma \sum_{j=1}^{120} (V_j - V_{j0})^2 + \sum_{i=1}^{2} (y_i - y_{p_i})^2, \tag{11}$$

where A denotes the resultant matrix on the left side of (7), and C is the vector consisting of known terms on the right side. The second term is the constraint for volume, where γ is the coefficient to adjust the tolerance of the volume changes of the tongue body. The volume tolerance was controlled within two percent in the present study. The third term is the jaw movement. The cost function is minimized by making its partial derivative with respect to x_i equal zero.

III. DYNAMIC ARTICULATORY CONTROL

There are two common strategies used for controlling a physiological model. One uses EMG signals as muscle activation

patterns to drive a physiological model [12, 14, 15]. Another approach controls a dynamical model of articulators based on the EPH, which claims that limb movements are produced by centrally specified shifts of the mechanical equilibrium of the peripheral motor system (namely, λ model) [22, 25]. The EPH is a plausible approach for controlling the dynamic movement of the tongue and other articulators, since the articulatory system is a network of muscles and floating rigid bodies. However, the most commonly used version of this theory requires length parameters and firing information of the muscles, which are difficult to obtain empirically. On the other hand, observation of EMG signals is limited to only a few large muscles such as the extrinsic tongue muscles [23]. For the above reasons, it is difficult to implement either the EPH approach or to use observed EMG signals to control a physiological model of articulation.

A. Construction of Muscle Workspace

In this study, our physiological model is driven by a muscle activation pattern (MAP), which consists of contraction signals for the muscles. The key point is how to develop an efficient approach to enable a mapping between the MAP and observable articulatory parameters. For this reason, we developed a practical control method that generates MAPs, which are consistent with EMG signals, according to articulatory targets.

The first step towards developing the target-based control strategy is to examine the effect of individual muscle contraction on tongue deformation. Figure 8 shows tongue deformations produced by exciting the four extrinsic tongue muscles with a 200-ms activation signal. The

Figure 8

activation of the GGp results in upward and forward movement of the tongue body, and the activation of the HG results in downward and backward movement, as shown in Fig. 8 (a) and (d). Likewise, the SG produces backward and upward movement of the tongue body, and GGa causes downward and forward movement with a midline groove formation. These simulation results are basically consistent with what were expected from the EMG data [4, 23]. If we define a representative portion such as the tongue dorsum as the observation point, the function of each muscle in the geometric space can be described by the relationship between the muscle activation level and the displacement of the observation point.

Supposing that only a single muscle is involved in the movement, we can estimate the muscle activation level from a displacement of the observation point, and vice versa. This suggests a general procedure for deriving muscle activation signals from an articulatory movement towards the given articulatory target for an observation point. For convenience, the observation point is hereafter referred to as a control point. When a single muscle is excited by a given activation signal in the model simulation, the control point moves from its initial position to a new position. This displacement forms a muscle force vector corresponding to the muscle contraction. The muscle force vectors can be obtained for all the muscles by independently exciting every muscle using a unit activation signal. All of the muscle vectors form a vector space for each control point. The vector space is referred to as a muscle workspace. The muscle workspaces reflect the relationship of the muscle activation and the articulatory movement.

B. Target-based Control Strategy

Since the muscle workspace is compatible with the geometrical space, the mapping of the control point between the geometrical space and muscle workspace is straightforward. Here, we use an example shown in Fig. 9 to explain the procedure of generating muscle activation signals according to a given target in a simplified muscle workspace. This muscle workspace consists of the muscle vectors of the four extrinsic tongue muscles, shown by the thick dark arrows. *Pc* indicates the current position of the control point and *Tg* is the target position. When the control point moves towards the target, the dashed line from *Pc* to *Tg* forms a vector, referred to as an articulatory vector. When the articulatory vector is mapped onto the muscle workspace, a set of projections is obtained for the muscle vectors. Supposing that the projection of the articulatory vector for muscle vector v_i is $\alpha_i v_i$ and the projection of the optimal vector of the control point moving towards the target is $\beta_i v_i$, a cost function is defined as the summation of the squared difference of each vector component between the articulatory vector and the optimal vector. By means of a penalty function, the component of the optimal vector can be solved by minimizing the cost function. As a result, the generated activation signal for the optimal vector is $\beta_i \approx u(\alpha_i)\alpha_i$ for muscle *i,* where $u(\alpha_i)$ is the unit step function, 1 for $\alpha_i > 0$, and 0 for the else. This means that the positive projections alone contribute to the movement toward the target, and the negative ones can be ignored. Thus, the SG and HG are the active muscles at the current computational step shown in Fig. 9. As the activation signals are computed at each computational step and fed to the muscles, the control

Figure 9

point is driven to approach its target. Figure 9 shows the resultant trajectory of the control point (thin gray path), where the gray arrow indicates the optimal vector at the current step. If there are multiple control points in a system, the resultant muscle activation signals are the summation of the signals for all control points.

Strictly speaking, the above process faces the inverse problem in deriving MAPs from displacements. When the muscle activation signal is derived from the distance between the control point and target, there may be more than one combination of the activation patterns that can drive the control point to move towards the target if we account for different levels of co-contraction among antagonistic muscles. This study, however, has not dealt with this problem, but assumed that there is no co-contraction between agonist-antagonist pairs.

IV. CONCLUSIONS

By adopting reasonable simplifications in modeling to compute the movement and deformation of the tongue tissue, we have developed a physiological articulatory model of speech organs, which is capable of simulating human speech articulation within an acceptable computational time. The major simplification involved that a lumped mass matrix is used to replace the consistent mass matrix, and a viscoelastic spring network is employed to approximate the non-zero elements in the matrices of viscous and stiffness components in the FEM-based governing motion equations. Based on this approach, both the soft tissue and rigid organs are modeled as a single mass-spring network. This mass-spring network showed a reliable performance in simulating a large and fast deformation of a soft tissue continuum, and the

computational time was significantly reduced compared with the other models [12].

A practical control strategy was developed to generate muscle activation patterns based on given targets, and then used to drive the physiological model. This approach relates articulatory movements of a control point to MAPs *via* a muscle workspace; it reduces the distance between the position of control points and the targets through a stepwise computation. The strategy is a reiterative procedure: the muscle activation signals are computed at each step based on the current position and the articulatory target, and are then fed to the muscles to drive the control point toward the articulatory target. The development of this control strategy started from the functional subdivisions in tongue movement that associate with disjoint subsets of the tongue muscles [18, 22], and arrived at a synergetic approach in which the control signals to all the tongue muscles contribute to the production of each of the basic motions [25].

The performance of the model was examined using the X-ray microbeam data obtained from the target speaker. The results showed that the model demonstrates dynamic characteristics that resemble the pattern of coarticulation in human speech articulation [5]. In this model, contact of the tongue tissue and rigid organs that takes place during the articulation was also treated in model articulations. This physiological articulatory model has been used in our ongoing studies to synthesize vowel-consonant-vowel sequences and short speech phrases, and to estimate the vocal tract shape from sound wave [6, 7].

REFERENCE

[1] J. Perkell, “A physiological-oriented model of tongue activity in speech production,” Ph. D. Thesis, MIT (1974).

[2] K. Bathe, *Finite element procedures*, Prentice-Hall, New Jersey (1996).

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

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

[5] J. Dang, and K. Honda, “Speech production of vowel sequences using a physiological articulatory model,” Proc. ICSLP98, Vol. **5**, 1767-1770 (1998).

[6] J. Dang, and K. Honda, "Speech synthesis of VCV sequences using a physiological articulatory model," J. Acoust. Soc. Am, **105**, p. 1091 (1999).

[7] J. Dang, and K. Honda, “Estimation of vocal tract shape from speech sounds via a physiological articulatory model”, 5th Speech Production Seminar, (Munich, Germany) (2000).

[8] E. L. DuBrul, *Sicher's Oral Anatomy*, Mosby, St.Louis (7th Edition) (1980).

[9] Y.C. Fung, *Biomechanics - Mechanical properties of living tissue*, Spriger-Verlag, New York (2nd Edition) (1993).

[10] K. Hashimoto, and S. Suga, “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 (1986).

[11] A. V. Hill, "The heat of shortening and the dynamic constants of muscle," Proc. Roy. Soc. London B **126**: 136-195 (1938).

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

[13] K. Honda, H. Hirai, and J. Dang, "A physiological model of speech organs and the implications of the tongue-larynx interaction," Proc. ICSLP94, 175-178, Yokohama (1994).

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

[15] Y. Kakita, O. Fujimura, and K. Honda, "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 (1985).

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

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

[18] S. Maeda, and K. Honda, "From EMG to formant patterns of vowels: the implication of vowel spaces," Phonetica, **51**, 17-29 (1994).

[19] K. Miyawaki, "A study of the muscular of the human tongue," Ann.

Bull. Res. Inst. Logoped. Phoniatrics Univ. Tokyo, **8**, 23-50 (1974).

[20] A. Morecki, "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, Spriger-Verlag, New York (1987).

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

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

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

[24] T. Sakamoto and Y. Saito, *Bionics and ME - From the basic to measurement control*, Tokyo Denki University Press, Tokyo (1980). (in Japanese)

[25] V. Sanguineti, J. Laboissière, and D. Ostry, "A dynamic biomechanical model for neural control of speech production," J. Acoust. Soc. Am. **103**, 1615-1627 (1998)..

[26] H. Takemoto, "Morphological Analyses of the Human Tongue Musculature for Three-dimensional Modeling," J. SLHR, 44, 95-107 (2001).

[27] J. Warfel, *The head, neck, and trunk*, Led & Febiger, Philadelphia

and London (1993).

[28] R. Wilhelms-Tricarico, “Physiological modeling of speech production: Methods for modeling soft-tissue articulators,” J. Acoust. Soc. Am. **97**, 3805-3898 (1995).

[29] K. Yamazaki, “The weight of the cranium and mandible with comparison of the dental and bony regions of the mandible,” Japanese Journal of Dentistry, 26: 769-796 (1933). (in Japanese)

[30] O. Zienkiewicz, and R. Taylor, *The finite element method,* McGraw-Hill Book Company, New York (1989).

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

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

Table 2 Organs and Muscles involved in the Model

Organs	Tongue, Hyoid bone, Jaw
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], Temporalis[c], Masseter, Stylohyoid, Sternohyoid, Stylopharyngeus

[o] The major muscles in the opener group;

[c] The major muscles in the closer group;

Figure Captions

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

Figure 2 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 3 The arrangement of the tongue muscles in the midsagittal and/or parasagittal planes. (b) and (c) show the arrangement in a parasagittal plane and the others for the midsagittal plane. (Dimensions in cm)

Figure 4 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.

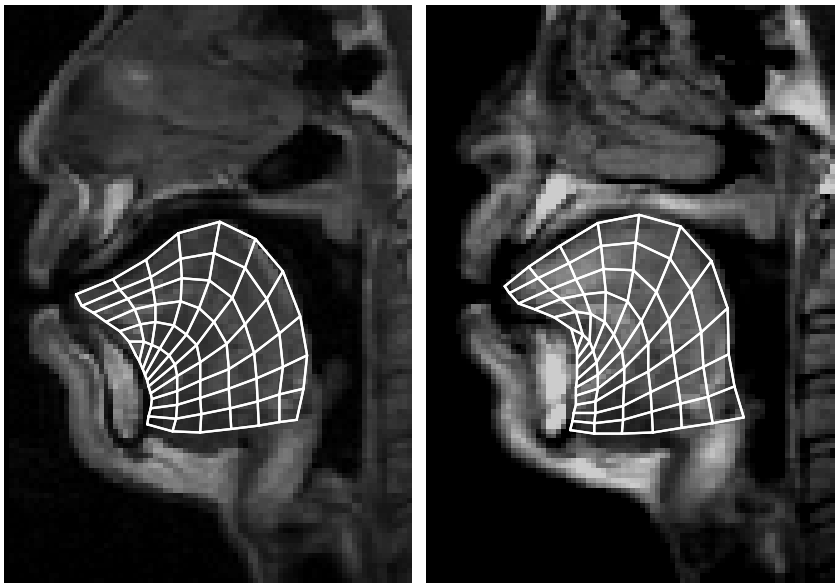
Figure 5 Modeling of the vocal tract wall: (a) extracted outlines of vocal tract wall based on MR images, and (b) reconstructed surface of vocal tract walls with the tongue body. (Dimensions in cm)

Figure 6 Muscle modeling: (a) a general model of muscle unit: k and b are stiffness and dashpot, E is contractile element. (b) generated force varies with stretch ratio ε. α is a gain of the active part to determine the maximum force.

Figure 7 A model of the articular groove of the tempromandibular joint and forces applied on the condyle. Force is redistributed at contact point of the condyle with the path. Dark lines show forces related to the x-direction and gray lines for the y-direction. f_x and f_y are initial inputs. $f_{x'}$ and $f_{y'}$ are redistributed forces.

Figure 8 Tongue deformations by the 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. The cross arrow indicates the direction of tongue movements by indicated muscles. (Dimensions cm)

Figure 9 An example of the proposed control strategy. O: initial position of the control point; *Pc*: current position, *Tg*: articulatory target. *fsg* and *fhg* are the positive projections of vector of *Pc* to *Tg*. Gray line shows trajectory from *Pc* to *Tg*, and gray arrow indicates the direction at the current step.



(a) (b)

Figure 1

J. Dang and K. Honda

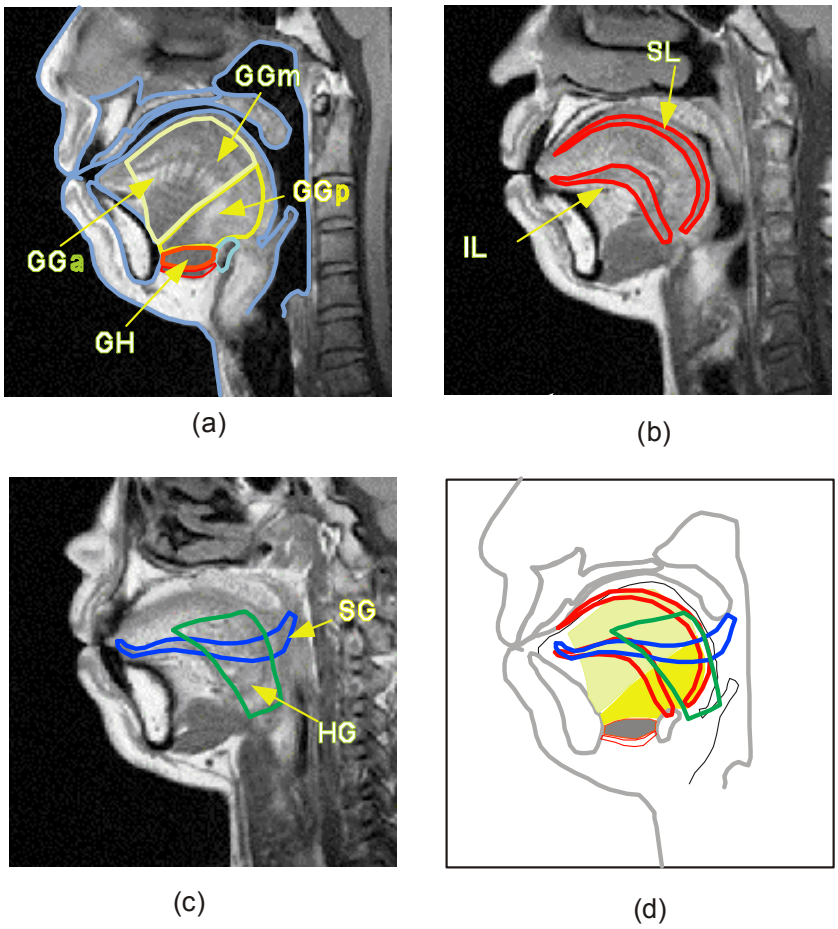


Figure 2

J. Dang and K. Honda

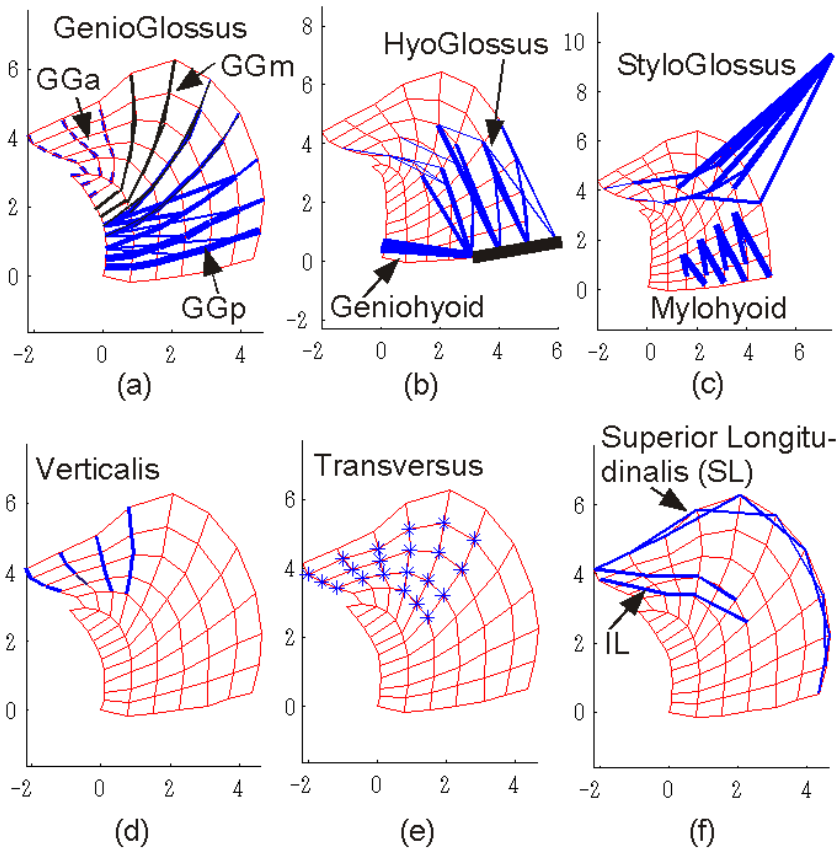


Figure 3

J. Dang and K. Honda

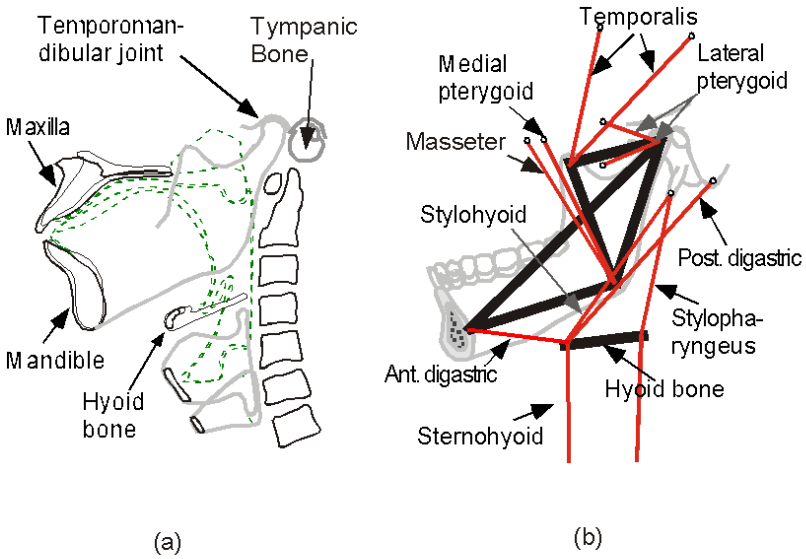


Figure 4

J. Dang and K. Honda

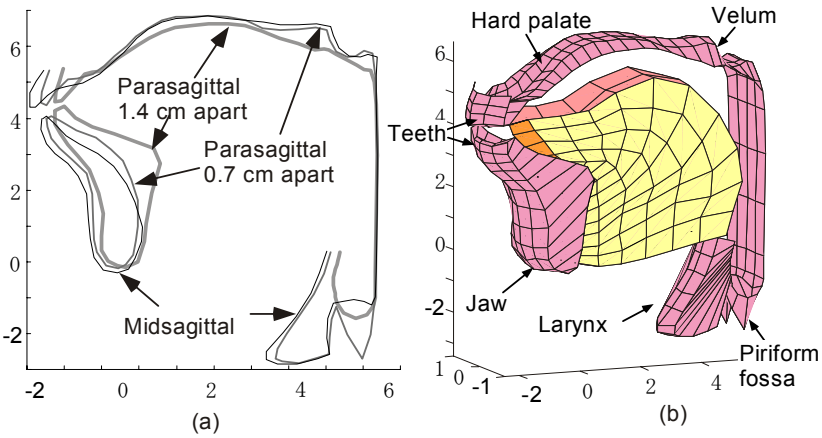


Figure 5

J. Dang and K. Honda

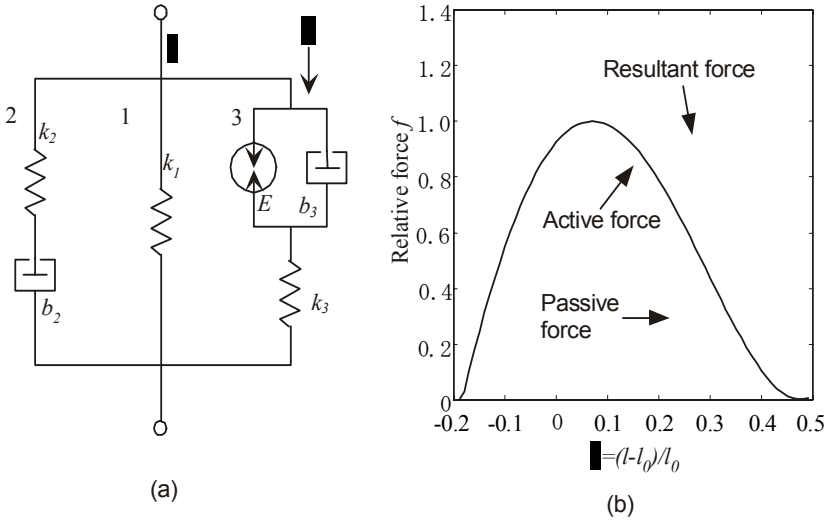


Figure 6

J. Dang and K. Honda

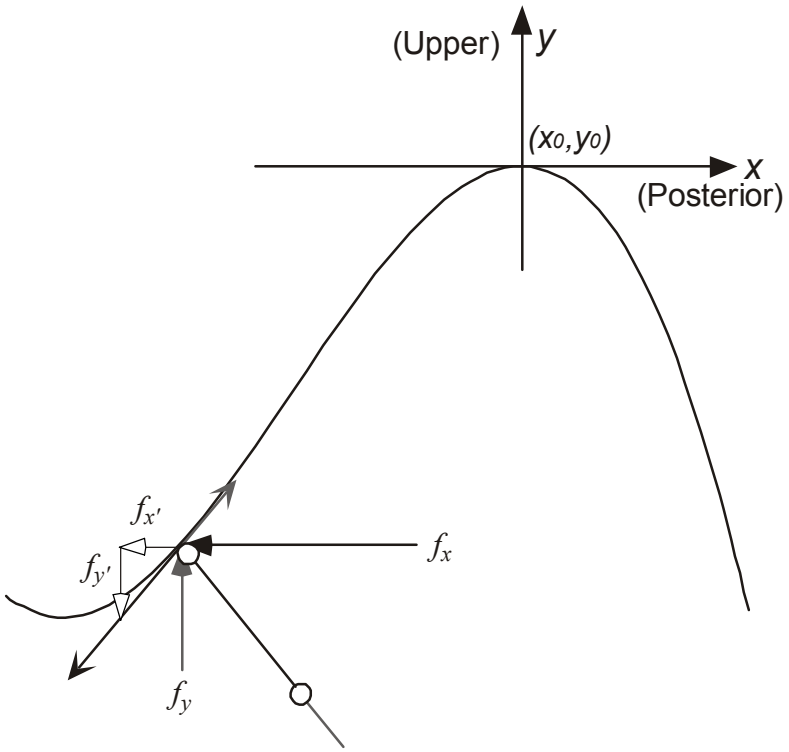


Figure 7

J. Dang and K. Honda

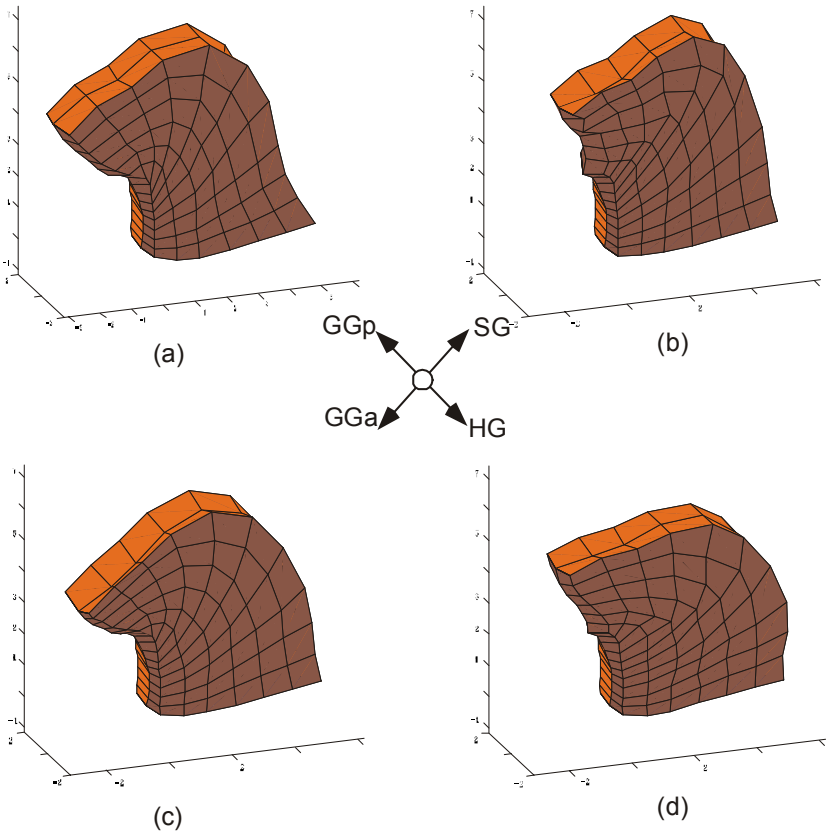


Figure 8

J. Dang and K. Honda

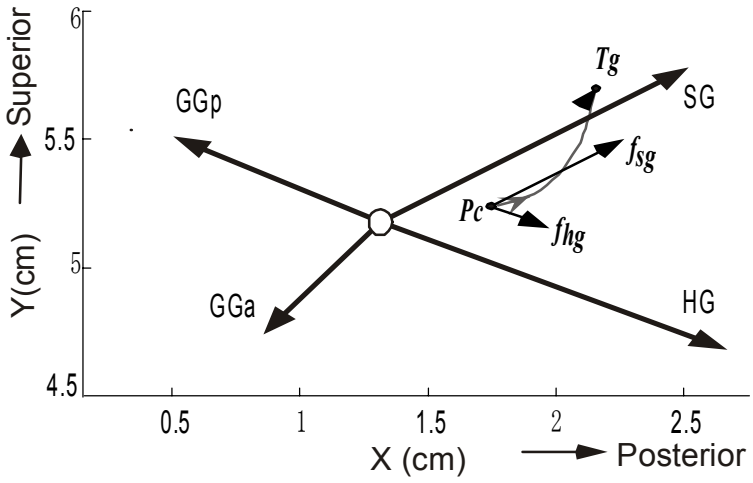


Figure 9

J. Dang and K. Honda