Construction and control of a physiological articulatory model

Jianwu Dang[a)]
School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan, and ATR Human Information Science Laboratories, Kyoto, Japan

Kiyoshi Honda
ATR Human Information Science Laboratories, Kyoto, Japan

(Received 16 May 2003; revised 17 October 2003; accepted 6 November 2003)

A physiological articulatory model has been constructed using a fast computation method, which replicates midsagittal regions of the speech organs to simulate articulatory movements during speech. This study aims to improve the accuracy of modeling by using the displacement-based finite-element method and to develop a new approach for controlling the model. A ''semicontinuum'' tongue tissue model was realized by a discrete truss structure with continuum viscoelastic cylinders. Contractile effects of the muscles were systemically examined based on model simulations. The results indicated that each muscle drives the tongue toward an equilibrium position (EP) corresponding to the magnitude of the activation forces. The EPs shifted monotonically as the activation force increased. The monotonic shift revealed a unique and invariant mapping, referred to as an *EP map*, between a spatial position of the articulators and the muscle forces. This study proposes a control method for the articulatory model based on the EP maps, in which co-contractions of agonist and antagonist muscles are taken into account. By utilizing the co-contraction, the tongue tip and tongue dorsum can be controlled to reach their targets independently. Model simulation showed that the co-contraction of agonist and antagonist muscles could increase the stability of a system in dynamic control.
[DOI: 10.1121/1.1639325]

PACS numbers: 43.70.Bk, 43.70.Aj, 43.70.Jt [AL] Pages: 853–870

I. INTRODUCTION

The production of speech involves the fine coordination of a serially ordered stream of changing articulatory, laryngeal, and respiratory movements. These skilled articulatory movements are concerned with well-practiced and overlearned muscular behaviors. A number of speech researchers have endeavored to discover the relationship between tongue movement and muscle activation via experimental approaches such as EMG experiments (Baer *et al.*, 1988; Dang and Honda, 1997). For the muscles involved in speech organs, however, experimental observations have succeeded for only a few large muscles, such as the extrinsic tongue muscles. The functions of the intrinsic muscles of the tongue have been investigated using tagged MRI (Niimi *et al.*, 1994), or tagged cine-MRI (Stone *et al.*, 2001), but the accuracy of such investigations is questionable, since a number of muscles are generally coactivated even to form a simple movement. Furthermore, it is often difficult to determine the mechanical load of a given muscle accurately because the load depends on the potential contribution of many other muscles. Therefore, a physiological model of the human speech organs is necessary for understanding the mechanism of speech production.

In the literature, a few physiologically based articulatory models have been reported. Perkell (1974) constructed the first computational model of the human tongue. His model was a two-dimensional (2D) projection of the tongue in the sagittal plane, which was composed of a lumped parameter and a lumped force system, equivalent to the finite-element method (FEM). The FEM approach has been applied to three-dimensional (3D) tongue models by Kiritani *et al.* (1976), Kakita *et al.* (1985), and Hashimoto and Suga (1986). These models were essentially based on infinitesimal elasticity methods, which describe the deformation process of the soft tissue as a sequence of quasistatic equilibrium configurations. Wilhelms-Tricarico (1995) proposed a method for modeling the tongue's soft tissue by large-scale FEM. Based on the method, he built a three-dimensional model of the tongue tissue and discussed the effects of geometric nonlinearities.

Hirai *et al.* (1995) developed a 2D physiological model unifying the tongue, jaw, and laryngeal structures (see also Honda *et al.*, 1994). The soft tissue of the tongue was formed by 2D FEM based on magnetic resonance imaging (MRI) data from a male speaker. The rigid organs such as the jaw, hyoid bone, thyroid cartilage, and cricoid cartilage were connected by muscles with elasticity. The dynamic balance of forces and moments was used as a mechanical principle to interface the soft and rigid structures. Because these structures were modeled independently, the model was computationally slow in achieving an equilibrium between the soft tissue and rigid bodies. Payan and Perrier (1997) reported a 2D biomechanical tongue model built by FEM. Their model produced V–V sequences according to the equilibrium point hypothesis (EPH), one of the common motor control theories (Feldman, 1986). The model demonstrated plausible movements for the tongue without incorporating jaw movement.

[a)]Electronic email: jdang@jaist.ac.jp

Sanguineti *et al.* (1998) employed a 2D model of the tongue, jaw, hyoid bone, and larynx to develop a control strategy based on the EPH (λ model). In their study, the dynamic behavior of the entire system was specified by its global kinetic and potential energy functions. They noted that the dynamic effects that occurred at the interface between the soft tissue and rigid organs were not negligible in modeling speech-like movements. This often resulted in serious stability errors because the dynamic processes of these two components were quite different.

In previous works (Dang and Honda, 2001; 2002), the authors measured the structure of speech organs based on volumetric MRI data obtained from a male speaker and constructed a partially 3D physiological articulatory model that consisted of the tongue, jaw, hyoid bone, and the vocal-tract wall. To obtain a time-efficient control for the physiological articulatory model, both the soft tissue and rigid organs were modeled as a mass–spring network that can efficiently and reliably simulate a large, fast deformation. A target-based control strategy was developed to generate muscle activation signals and realize the model's dynamic articulatory movements. The model was used as a synthesizer to generate speech sounds of short phrases (Dang and Honda, 1998; Dang *et al.*, 1999). It was later also used to estimate vocal-tract shapes from speech sounds (Dang and Honda, 2002). The model's disadvantage was that one of the important physical parameters for an elastic continuum, the Poisson ratio, was not taken into account, since the framework of the model consisted of mass points and volumeless springs. The revised model described in this paper has a marked improvement: the use of a viscoelastic cylinder to replace the volumeless spring. This enhancement accounts for the Poisson ratio and thus achieves a ''semicontinuum'' tissue model for the tongue body. In the present study, the new model is used to examine muscle functions for the tongue based on the model simulations. Furthermore, a control method is proposed to handle co-contraction during speech production.

II. MODELING USING DISPLACEMENT-BASED FINITE-ELEMENT METHOD

During natural speech, the tongue forms lateral airways by narrowing the tongue blade, or it makes a midsagittal conduit by grooving with bilateral tongue–palate contact, as seen in some consonants (e.g., /s/) and some vowels (e.g., /i/). A realistic model of the tongue must be capable of forming such midsagittal conduits and side airways, which are essential behaviors of the tongue in speech production. As a trade-off between computational cost and model verisimilitude, we have constructed a partial 3D model with a thick sagittal layer instead of a full 3D model.

A. Configuration of physiological articulatory model

The essential configuration of the model is the same as that used in the previous studies (Dang and Honda, 2001; 2002). The tongue model is a partial sagittal representation of a volumetric MR image of the tongue, which was obtained from a male Japanese speaker. According to our observations (Dang *et al.*, 1997), the tongue contacts the hard palate in the lateral area 1.5 cm from the midsagittal plane during most of the phonations. Such a contact interferes with the representation of the inherent characteristics of the tongue biomechanics. To achieve accurate modeling, the layer on each side of the midsagittal plane should be thinner than 1.5 cm in a model with two symmetric layers such as the proposed model of this study. Therefore, the lateral bound of the layers was set to be 1.0 cm apart from the midsagittal plane. The outlines of the tongue body are 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. Under the assumption that the left and right sides of the tongue are symmetrical, the outline of the left side is copied to the right. The initial shape of the model adopts the tongue shape of a Japanese vowel [e], which approximates a centralized vowel in Japanese. Mesh segmentation of the tongue tissue roughly replicates the fiber orientation of the genioglossus muscle, the largest muscle in the tongue. The outline of the tongue body in each plane is divided into ten radial sections that fan out from the attachment of the genioglossus on the jaw to the tongue surface. In the perpendicular direction, the tongue tissue is divided concentrically into six sections. A 3D mesh model is constructed by connecting the section nodes in the midsagittal plane to the corresponding nodes in the left and right planes, where each mesh is a ''brick'' with eight corner nodes. Thus, the model represents the principal region of the tongue as a 2-cm-thick layer bounded by three sagittal planes. Figure 1 shows the initial shape of the tongue model based on the segmentation, with the surrounding organs. In this segmentation, the tongue tissue is represented as 120 eight-cornered brick meshes.

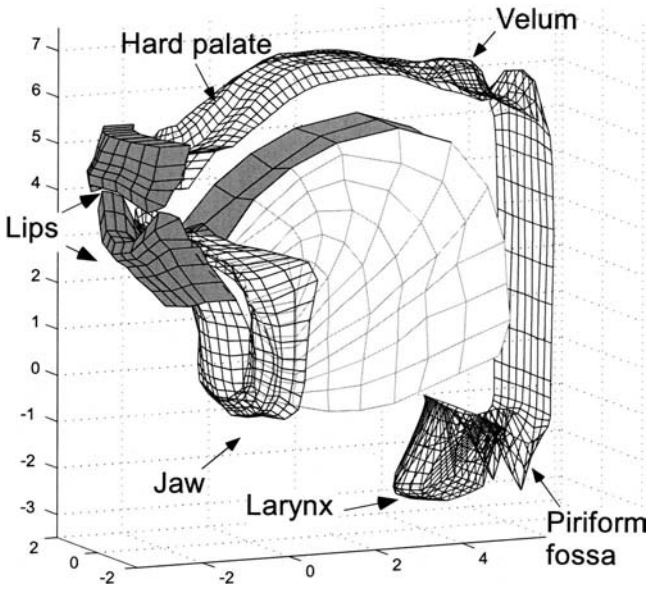

FIG. 1. Configuration of the physiological articulatory model.

To generate a vocal-tract shape, the articulatory model must include the tongue, lips, teeth, hard palate, soft palate (the velum), pharyngeal wall, and larynx. At the present stage, the lips and the velum are not modeled physiologically. They are included in the construction of vocal-tract shapes for speech synthesis but not in the generation of articulatory movements. The lips are defined by a short tube with a length and cross-sectional area, and the movement of

the velum is described by the opening area of the nasopharyngeal port. Outlines of the vocal-tract wall and the mandibular symphysis were extracted from MR images in the midsagittal and parasagittal planes (0.7 and 1.4 cm from the midsagittal plane on the left side). Again, assuming that the left and right sides are symmetrical, 3D surface models of the vocal-tract wall and the mandibular symphysis were constructed using mesh outlines with 0.7-cm intervals in the left–right direction. Figure 1 shows the model configuration of the vocal tract.

B. Modeling the tongue tissue using a truss structure

The soft tissue of the tongue has been commonly modeled using the finite-element method (FEM) (Kakita *et al.*, 1985; Wilhelms-Tricarico, 1995). The present study adopts displacement-based FEM as the basis of the modeling effort, which is referred to as an extended finite element method (X-FEM). The principal advantages of X-FEM are that the finite-element framework (sparsity and symmetry of the stiffness matrix) is retained and that a single-field variational principle is used (Zienkiewicz and Taylor, 1989; Belytschko *et al.*, 2002). Based on the principle of this X-FEM (see the Appendix for details), we can obtain the motion equation of equilibrium (1) for governing the linear dynamic response of a finite-element system

$$M\ddot{X}+B\dot{X}+KX=F, \tag{1}$$

where X, $\dot{X}$, and $\ddot{X}$ are the displacement, velocity, and acceleration vectors of the finite element assemblage, respectively. M, B, and K are the mass, damping, and stiffness matrices, respectively, in which the description is based on the displacements of the nodal points. F is the vector of externally applied loads.

1. Displacement function

The essence of X-FEM is that it uses the node displacements to describe the 3D deformation of a continuum. Exploiting this advantage, we reconsider the brick mesh mentioned above and then reduce some of the complexities in modeling the tongue tissue based on the common assumption that tongue tissue is an isotropic material.

In the model configuration, the basic mesh unit is described as an eight-node brick. Focusing mainly on the displacement of those nodal points, the relation of the relative locations of the nodes becomes a key point for analysis. For an isotropic material, this relation between the nodes can be easily represented by using an elastic solid to connect the nodes in all directions within the eight-node mesh. Based on this consideration, we use Hookean elastic bodies (Fung, 1993), referred to hereafter as cylinders, to connect all of the adjacent nodal points of the eight nodes in 3D. In the hexahedral mesh, there are 12 edges, corresponding to 12 cylinders. On the surface, there are two cylinders connecting the diagonal nodes crosswise on each of the six planes. Inside the brick, there are four cylinders transversely connecting the diagonal vertices. Altogether, each eight-node mesh is constructed from 28 viscoelastic cylinders. If the relative movement of every node pair can be described correctly by the changes in the cylinders, the node displacement in 3D can be fully described. Thus, the basic unit for the X-FEM analysis degenerates from the brick mesh to a truss. The displacement function for each truss is simplified as

$$H=\frac{1}{l_0}\begin{bmatrix} l_0-x & 0 & 0 & x & 0 & 0 \\ 0 & l_0-x & 0 & 0 & x & 0 \\ 0 & 0 & l_0-x & 0 & 0 & x \end{bmatrix}\begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix}, \tag{2}$$

where x is the distance from node i to node j, and l_0 is the length of the cylinder. T is a transformation matrix from a local coordinate (x, y, z) to the global coordinate (X, Y, Z)

$$T=\begin{bmatrix} \cos(\mathrm{X,x}) & \cos(\mathrm{X,y}) & \cos(\mathrm{X,z}) \\ \cos(\mathrm{Y,x}) & \cos(\mathrm{Y,y}) & \cos(\mathrm{Y,z}) \\ \cos(\mathrm{Z,x}) & \cos(\mathrm{Z,y}) & \cos(\mathrm{Z,z}) \end{bmatrix}. \tag{3}$$

Since each cylinder is a *Hookean elastic solid*, its deformation obeys Hooke's law. For an isotropic material, the stress–strain matrix is a 6×6 matrix described by Young's modulus E and Poisson ratio ν. Supposing that the Hookean body is a uniform cylinder with a radius r, its cross-sectional area varies uniformly when a force is loaded in the axial direction (x direction) alone. Thus, the 3D relation between the cylinder's length and its perpendicular dimension is simplified to

$$r=r_0\left(1-v\,\frac{l-l_0}{l_0}\right), \tag{4}$$

where l_0 and l are the lengths of the cylinder before and after applying a force and r_0 and r are the cylinder's radii corresponding to these lengths. At the end, the stress–strain matrix degenerates into a constant. When an axial force is loaded, the length variation of the elastic cylinder depends on the Young's modulus, while the change in the thickness obeys the law shown in (4). The deformation of a cylinder is transformed into the global coordinate by (2). Displacements of the nodes in a brick mesh are thus described by the deformation of the cylinders in the truss structure. Note that in the tissue model, each node connects seven cylinders (at the corners of the model) to 26 cylinders (inside the model). When a force is applied to a node, it is decomposed into seven to 26 directions corresponding to the axial direction of the connecting cylinders. Using the force decomposition, any force transmission within the model is achieved by a set of axial forces of the concerned cylinders. Therefore, we would not expect the simplification to introduce any significant error.

The dissipation caused by velocity-dependent damping must be taken into account when simulating the dynamic responses of the speech organs. For this purpose, we treat the cylinder as a viscoelastic body. According to Fung (1993), there are three types of models for representing a viscoelastic material: the Voigt model, the Maxwell model, and the Kelvin model. All three are composed of combinations of linear springs and a dashpot. The Voigt model consists of a spring parallel to a dashpot, the Maxwell model consists of a spring cascaded to a dashpot, and the Kelvin model is a combination of the two models. The relation of force F, displacement x and velocity $\dot{x}$ is described in (5a) for the Voigt model and (5b) for the Maxwell model

$$F = kx + b\dot{x}, \tag{5a}$$

$$x = \dot{F}/k + F/b, \tag{5b}$$

where k and b denote the stiffness and viscous coefficients, respectively. The Voigt model is good for describing a solid body, while the Maxwell model is good for liquid material. Therefore, we have adopted the Voigt model in modeling the tongue tissue. Comparing (5) and (1), one can also see that the Voigt model is easy to incorporate into the motion equation. When a force is applied on the Voigt model, a deformation gradually builds up as the spring shares the load. After the force is released, the dashpot displacement relaxes exponentially, and the original length is restored from the deformation.

2. Volume of cylinders

As described above, a hexahedral mesh with eight nodes consists of 28 cylinders. The tongue tissue is assembled from 120 such hexahedral meshes. The volume of the cylinders is determined by the basic principle that the volume summation of the cylinders concerning a mesh must be equal to the volume of the mesh.

In the assembled meshes, a cylinder can be shared by several adjacent meshes. In this case, the volume of the cylinder is assumed to distribute equally over the shared meshes. Thus, the weight coefficient of a cylinder for the relevant meshes is equal to the reciprocal of the number of the meshes sharing the cylinder. Among the 28 cylinders, there are four cylinders connecting the diagonal nodes inside a brick mesh, which are concerned with this mesh alone. Therefore, their weight coefficients are 1.0 in calculating the volume. The weight coefficient for the surface cylinders is 0.5 because they are shared by two adjacent meshes. The weight coefficient is 0.25 for the edge cylinders that are shared by four adjacent meshes.

The length of each cylinder is calculated in terms of the nodal coordinates of the two ends. To calculate the cross-sectional area of the cylinders, the shape of a mesh is first imagined as a uniform cylinder, whose length is the summation of the weighted length of all cylinders of this mesh. Then, the cross-sectional area of the uniform cylinder is determined by the quotient of the mesh's volume to the equivalent length. As a result, the thickness of the inside cylinders equals the cross-sectional area. The thickness for the surface cylinders is the summation of half the cross-sectional areas of two sharing meshes. Similarly, the thickness of the edge cylinders is the summation of the weighted cross-sectional areas of four cylinder-sharing meshes.

In this model, the viscoelastic cylinder is the basic element. The cylinder is a continuum that fully obeys physical laws. To reduce the effects of the discreteness of the truss-structure meshes, the Poisson ratio is also taken into account in the 3D meshes via a volume constraint that is described in the following section. Thus, this model can be thought of as a semicontinuum. This is one of the primary improvements from the previous version, where the meshes consisted of mass points and volumeless springs (Dang and Honda, 2001; 2002). With this improvement, the mass matrix becomes a consistent matrix, instead of the lumped mass matrix used in the previous version.

TABLE I. Parameters used in the present model.

Tongue Tissue	Density 1.0 g/cm^3	Young's modulus 20 kPa	Viscosity 2.0 kPa·s	Poisson ratio 0.49
Mandible	Weight 150 g	Young's modulus 9.6×10^6 kPa	Viscosity 9.6×10^5 kPa·s	⋯ ⋯
Hyoid Bone	Weight 5 g	Young's modulus 9.6×10^6 kPa	Viscosity 9.6×10^5 kPa·s	⋯ ⋯

3. Parameters and testing

In this study, the cylinder element is treated as a viscoelastic body. To describe the elastic properties, two essential parameters, Young's modulus E and Poisson ratio ν, are employed. A damping property parameter b is also introduced to describe the viscous property.

The Young's modulus influences the deformation of the body in the direction of the force, which basically measures the stiffness of the material. However, such mechanical parameters reported in past studies, which were obtained from different parts of the human body or from animals, have differed widely. Among them, Oka (1974) reported a value of 300 kPa for a contracted muscle. Duck (1990) found a value of 6.2 kPa for a human muscle under the rest condition, and a value of 110 kPa for the same muscle when contracted. Min *et al.* (1994) reported a Young's modulus of 20 kPa for the soft tissue of the vocal folds. Payan and Perrier (1997) used 15 kPa for the tongue tissue in their modeling study, and values ranging from 15 to 250 kPa for the tongue muscles corresponding to different levels of contraction. In this study, a value of 20 kPa is used for the Young's modulus in modeling the tongue tissue, following Min *et al.* (1994). For the viscosity, the viscous coefficient was 2 kPa·s, which was determined by a numerical experiment using this model. This value is about one-tenth that of the Young's modulus. Poisson ratio ν was set to 0.49, which is similar to that used in previous studies (Wilhelms-Tricarico, 1995; Payan and Perrier, 1997). The soft tissue of the tongue is considered to possess the same density as water. Therefore, a value of 1.0 g/cm^3 was used for the density of tongue tissue. These parameters are listed in Table I.

To verify the behaviors of the semicontinuum model with the Poisson ratio, we applied a force on a cuboid built on the truss structure and compared the Poisson ratio estimated from deformation of the cuboid with the original one. It was found that the estimated Poisson ratio varies with applied forces. To reduce this artifact, one more parameter, the ratio of the stress to the longitudinal strain, was employed. Duck (1990) showed that the Young's modulus depends to some extent on the ratio of the stress to the elongation, and is nearly constant for the soft tissue of the muscle if the relative elongation is less than 20%. The Young's modulus exponentially increases when the elongation ratio is larger than 20%. Following Duck, the Young's modulus is treated as a constant of 20 kPa in this study when the elon-

gation ratio is less than 20%, and its value increases as the elongation ratio becomes larger. The increasing rate of the Young's modulus functions as a parameter to maintain a constant Poisson ratio during tissue deformation.

The elongation-dependent Young's modulus was determined using a viscoelastic cuboid with a size of $6\times3\times2$ cm^3. The cuboid was divided into $5\times3\times2$ hexahedral meshes and constructed using the truss structure. When applying forces on the cuboid in the longitudinal direction, the Poisson ratio was evaluated by the changes of the cuboid in the axial and perpendicular directions using formula (4). An elongation-dependent Young's modulus was chosen to maintain the Poisson ratio around 0.49 while treating the cuboid either as a brick-assembled body or as a single brick. A numerical simulation was also conducted on the same cuboid by using ANSYS™ software, and the same behaviors were confirmed. The simulation showed that the truss-structure model based on X-FEM is sufficient for modeling the tongue tissue.

Compared with the traditional FEM, there are two major benefits of using this modeling. First, a fast computation is achieved without significant loss in accuracy. Second, the proposed model demonstrates excellent stability; no divergence was seen, even when a quite large force was loaded or when extreme deformation took place.

4. Volume constraint for tongue body

The tongue body is commonly considered to consist of incompressible tissue. However, the cubic meshes lack incompressible properties in the above model. If no volume constraint is taken into account, changes in the volume of the tongue were about 5% during tongue deformation. To reduce the changes, it is necessary to incorporate a constraint to maintain the volume of the tongue tissue. For this purpose, the Lagrange function first comes to mind. However, the constraint of the volume constancy introduced by the Lagrange function did not always work well; it sometimes interfered with tongue movement. One of the resulting phenomena, for example, was that on occasion the tongue could not move in response to a small change in force. It might be that the vectors for retaining constant volume are not distributed continuously over the multidimensional space consisting of the nodes' coordinates. For this reason, a procedure for minimizing volume changes is introduced to reproduce tissue incompressibility.

In the truss-structure model, the total volume of the tongue equals the summation of the volumes of all cylinders. Therefore, minimizing the changes in volume for all cylinders can achieve a volume constraint on the tongue body. When a force is applied on cylinder i in the axial direction, the change in the cylinder's volume is

$$\Delta V_i(X) = \pi r_0^2 l_0 - \pi r_0^2 \left(1-\nu\frac{\Delta l}{l_0}\right)^2 (l_0+\Delta l), \tag{6}$$

where the variation of the radius is represented by length increment Δl and Poisson ratio ν using (4). Using the Houbolt integration method (Bathe, 1996), motion equation (1) is rewritten in the finite difference expansions of $DX=B$, where D denotes the resultant matrix and B is the vector consisting of known terms in (1). The constraint is combined with the motion equation system by adding the volume difference and then minimizing the total error. Thus, the final system equations are derived from the following formula:

$$\frac{\partial}{\partial X}\left[\|DX-B\|^2+\alpha\sum_i \Delta V_i(X)^2\right]=0, \tag{7}$$

where α is the coefficient to adjust the tolerance of the volume changes in the tongue body. After introducing the volume constraint, the variation ratio of the volume of the tongue is reduced from about 5% to about 0.3%.

III. MUSCULAR STRUCTURE AND FORCE GENERATION

This model involves three kinds of external forces: muscle contraction, collision of the soft tissue with the rigid boundaries, and gravitational force. Muscular contraction is the source force to drive the model. Gravitational force always acts on all the nodes of the model in the vertical direction. The acceleration of gravity used in this model was 980 dyne·cm/s^2.

A. Modeling of muscles and rigid organs

The anatomical arrangement of the major tongue muscles was examined based on a set of high-resolution MR images obtained from the prototype speaker (Dang and Honda, 2001; 2002). The boundaries of the muscles were first traced in each slice of the MR images, and then superimposed on each other so that the contours for the major muscles could be identified. Thus, the genioglossus (GG) and geniohyoid (GH) were identified in the midsagittal plane, while the hyoglossus (HG) and styloglossus (SG) were mainly found in the parasagittal planes. The superior longitudinalis (SL) and inferior longitudinalis (IL) muscles were seen in both the midsagittal and parasagittal planes. The other intrinsic muscles such as the transversus and verticalis could not be identified in the MR images. The orientation of the tongue muscles was also examined with reference to the literature (Miyawaki, 1974; Warfel, 1993; Takemoto, 2001).

Figure 2 shows the arrangement of the tongue muscles used in the proposed model. Figure 2(a) shows the GG, which runs midsagittally in the central part of the tongue. 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) indicated by the dashed lines, the middle portion (GGm) shown by the gray lines, and the posterior portion (GGp) indicated by the dark lines. The thickness of the lines represents the approximate size of the muscle fibers: the thicker the line, the larger the maximum force generated. Figures 2(b) and (c) show the arrangement of other extrinsic muscles, the HG and SG, in the parasagittal plane, where the thickest line represents the hyoid bone. In addition, two tongue-floor muscles, the geniohyoid and mylohyoid, are also shown in the parasagittal planes. The top points of the mylohyoid bundles are attached to the medial surface of the

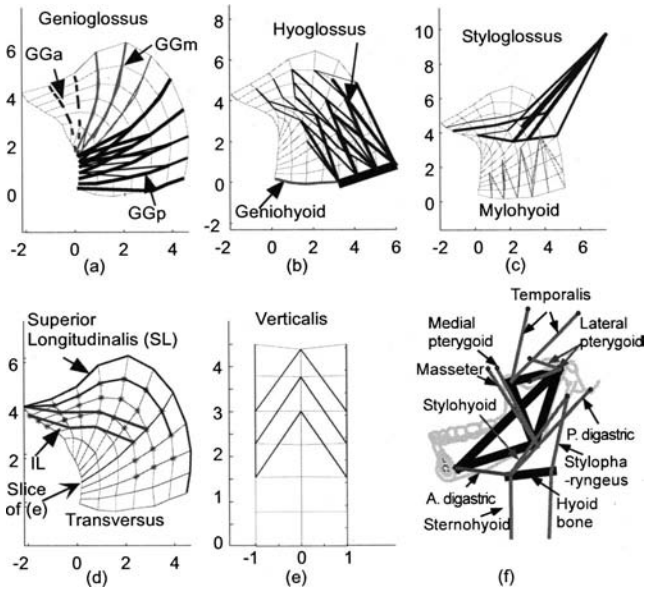

FIG. 2. Muscular structure of the model. (a)–(e) The tongue muscles in the midsagittal and/or parasagittal planes (dimensions in cm). (f) The mandible and hyoid bone complex.

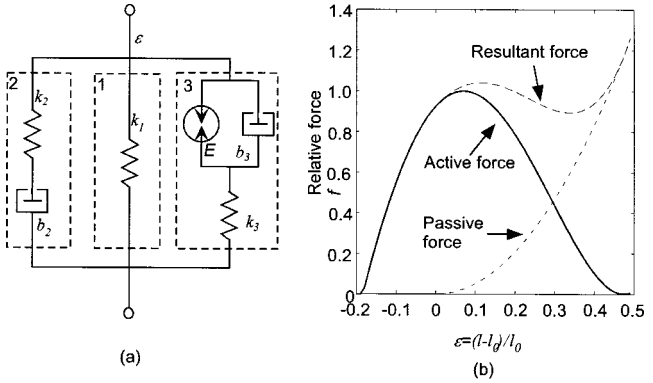

FIG. 3. Muscle modeling: (a) a general model of muscle unit: k and b are stiffness and dashpot, E is the contractile element; (b) generated force varies with stretch ratio ε.

mandibular body. All of the muscles are designed symmetrically on the left and right sides. Figure 2(d) shows the three intrinsic muscles of the SL, IL, and transversus. The transversus runs in the left–right direction, and its distribution is plotted in star markers. Figure 2(e) shows the structure of the verticalis muscle in a cross-sectional view sliced at the 5th section from the tongue floor, shown in Fig. 2(d). Altogether, 11 muscles are included in the tongue model.

Figure 2(f) shows the model of the jaw–hyoid bone complex. The right half of the mandible is drawn in the background with the pale gray lines. The model of the jaw has four nodes on each side, which are connected by five rigid beams (thick lines) to form two triangles with a shearing beam. These four points, which are similar to those used by Laboissière *et al.* (1996), are selected as the attachment points for the jaw muscles. The jaw model is combined with the tongue model at the mandibular symphysis. The hyoid bone is modeled as three segments corresponding to the body and bilateral greater horns. Each segment has two nodes connected with a rigid beam. Eight muscles indicated by thin lines are incorporated in the model of the jaw–hyoid bone complex, where the structure of the muscles is based on the anatomical literature (Warfel, 1993). The small circles indicate the fixed attachment points of the muscles. Since the 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 digastric muscle has two bellies, anterior and posterior, and is modeled to connect the hyoid bone at a fixed point. All of these muscles are modeled symmetrically left and right. Jaw movements in the sagittal plane involve a combination of rotation (change in orientation) and translation (change in position). Although there is no one-to-one mapping between muscle actions and kinematic degrees of freedom, 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.

According to Duck (1990), the Young's modulus of the bones is $11.5-12.0\times10^6$ kPa for the human femur and 6.9×10^6 kPa for the human tibia. In this study, we used the value obtained from the human tibia in modeling the bony structures, the mandible and the hyoid bone. For the weight of the bones, Yamazaki (1933) 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 90 g. According to this literature, the equivalent mass of the living jaw is roughly estimated to be 150 g including 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. The masses are uniformly distributed over the hard beams. To provide a uniform computational format, rigid beams were also treated as viscoelastic links with a high Young's modulus so that they could be integrated with the soft tissue in the motion equation. The viscosity was set to be about one-tenth of the Young's modulus in value, the same ratio as that used for the soft tissue. Note that the viscosity for the rigid beams is less important because the muscle force is not strong enough to cause any definite deformation on the rigid beams.

B. Generation of muscle forces

In formulating a generalized model of the muscle, this study adopts a commonly accepted assumption: a force depending on muscle length is the sum of the passive component (independent of muscle activation) and the active component (dependent on muscle activation). Figure 3(a) shows a diagram of the rheological model for a muscle sarcomere (Morecki, 1987), which is an extended model of Hill's model (Hill, 1938). The muscle model consists of three parts that describes the nonlinear property, the dynamic (force-velocity) property, and the force–length property. The properties of the muscle sarcomere can be described by a set of differential equations

$$\sigma_1 = k_1 \varepsilon,$$

$$\frac{\dot{\sigma}_2}{k_2} + \frac{\sigma_2}{b_2} = \dot{\varepsilon},$$

$$(\sigma_m + \sigma_3)(k_3 + E) + \frac{k_3 d(\sigma_m + \sigma_3)}{dt} = b_3 E \dot{\varepsilon} + E k_3 \varepsilon, \tag{8}$$

$$\sigma = \sigma_1 + \sigma_2 + \sigma_3,$$

where σ_1, σ_2, σ_3 are the stresses of each part, and σ is the total stress of the sarcomere; $\sigma_m = E\varepsilon^2$, and $\varepsilon = (l - l_0)/l_0$. l is the current length of the muscle sarcomere, and l_0 is the original length of the muscle sarcomere at rest position of the tongue.

The first three equations in (8) describe parts 1, 2, and 3 of the muscle model. Part 1 is a nonlinear spring k_1, which is involved in generating force only when the current length of the muscle sarcomere is longer than its original length. The value of k_1 is selected as $k_1 = 0.05 k_0 \varepsilon$, where $\varepsilon > 0$ and k_0 is the stiffness of the tongue tissue. Part 2 consists of a Maxwell body and is always involved in force generation. According to Eq. (5b), the force generated by this part is determined by two factors: the velocity of the muscle length and the previous force of this branch. As shown in the literature (cf. Zajac, 1989; Wilhelms-Tricarico, 1995; Laboissière *et al.*, 1996), the force–velocity characteristic of the muscle is treated as independent of the previous force. To emphasize the effect of the velocity of the muscle length, a relatively larger stiffness and a smaller viscous component are used in this part. The value of k_2 was set to be twice that of the tongue tissue, while b_2 was on the order of one-tenth that used in the tongue body.

Part 3 of the muscle sarcomere corresponds to the active component of the muscle force, which is the Hill's model consisting of a contractile element parallel to a dashpot and then cascaded with a spring. This part generates force as a muscle is activated; its characteristics are described by the third equation. In model computations, however, we use a force–length function of the muscle tissue instead of the third equation. The force–length function was derived by matching the simulation and empirical data using the least-square method (Morecki, 1987). The function arrived at a fourth-order polynomial of the stretch ratio of the muscles

$$\sigma_3 = 22.5\varepsilon^4 + 3.498\varepsilon^3 - 14.718\varepsilon^2 + 1.98\varepsilon + 0.858, \tag{9}$$

which has a similar shape to that used by Wilhelms-Tricarico (1995). This empirical formula is valid for $-0.185 < \varepsilon < 0.49$. The active force is assumed to be zero if ε is out of the given range. Figure 3(b) shows the relationship between the stretch ratio of the muscle sarcomere and the generated force including the passive force. This figure demonstrates the force–length characteristic of the muscle model.

Since a muscle consists of a number of muscular fibers with various lengths and thicknesses, the general lumped rheological parameters of the muscle tissue are not sufficient for determining the muscle-generated force. For this reason, we introduced a parameter, the ''thickness'' of the muscle fiber, into the force generation. The thickness works as a coefficient for all three parts of the muscle sarcomere, which ranged from 0.1 to 4. The value for a given muscle is determined by making the maximum force of the muscles consistent with empirical data (Laboissière *et al.*, 1996; Sanguineti *et al.*, 1997). As shown in Fig. 2, for example, the GGp is thicker than the GGa; so the GGp generates a stronger force than the GGa.

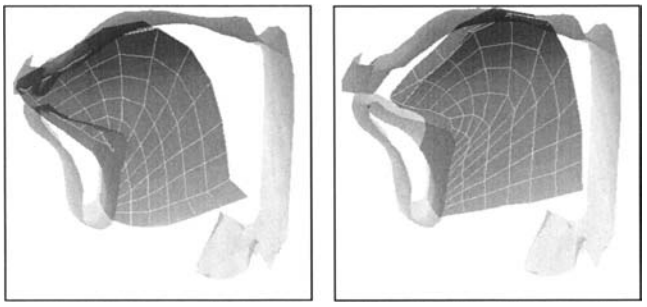

FIG. 4. Tongue deformation and movement from the initial position 20 ms after applying forces on GGp (left) and SG (right), where the jaw-opener muscle was activated in both cases. The force on the tongue muscles was 4 N, and 3 N for the jaw-opener muscle.

C. Evaluation of the articulatory model

Having completed the construction of the proposed articulatory model, including the soft tissue, rigid organs, and muscular structure, we conducted several numerical experiments to evaluate the characteristics of the model.

Figure 4 shows examples of the movements of the tongue and jaw driven by a force of 4 N for a tongue muscle and 3 N for the jaw-opener muscles, where the computation step was 5 ms in this simulation. The left panel demonstrates the tongue shape derived from the rest posture 20 ms after force was applied on the GGp and the jaw opener. The tongue tip and tongue dorsum moved upward and forward about 0.5 and 0.3 cm, respectively. The jaw lowered 0.16 cm. In most cases, tongue movement had a high positive correlation with jaw movement. An opposite movement was set for the tongue and jaw in order to test some of the extreme articulations that might occur in emphasized speech (Erickson, 2002). When the tongue and jaw moved in the opposite directions, the bilateral sides of the tongue blade contacted with the hard palate within 20 ms. The collision of the tongue and surrounding wall was not considered in this testing.

The right panel shows the posture of the model derived from the rest position 20 ms after applying a force of 4 N on the SG and 3 N on the jaw opener. Displacements were about 0.16 cm for the jaw and 0.8 and 0.6 cm for the tongue tip and tongue dorsum, respectively. The corresponding velocity for the control points was 41 cm/s for the tongue tip, 31 cm/s for the dorsum, and 8 cm/s for the jaw. For comparison, the authors measured the maximum velocity of the articulators from the articulographic data obtained from three speakers (Okadome and Honda, 2001). The maximum speeds were 40 cm/s for the tongue tip, 32 cm/s for the dorsum, and 12 cm/s for the jaw. The velocity was about the same for both the tongue tip and the dorsum in the measurement and model simulation, while the velocity of the jaw in this example was smaller than that of the measurement. A similar measurement

TABLE II. Comparison of the velocities (cm/s) of the tongue tip, tongue dorsum, and mandible obtained from simulation and observations.

	Tongue tip	Tongue dorsum	Mandible
Model	41	30 (18)[b]	8
EMMA	40	32 (22)[b]	12
Stevens (2000)[a]	32	15[b]	⋯

[a]Measured from the figures (Stevens, 2000).
[b]Vertical movement alone.

in the literature (Stevens, 2000) is cited in Table II and was used to evaluate the model simulation. The comparison showed that our simulation of the velocities was consistent with those results for the tongue tip and the tongue dorsum.

In Fig. 4, one can see that the meshes in the tongue demonstrated well-balanced deformation. In the right panel, the tongue dorsum made a closure with the palate within 20 ms, and the closure was maintained well until the jaw reached its stationary position with an approximately 1.0-cm aperture. The same test was also conducted in some extreme cases of applying a large force of 7 N on a number of tongue muscles. For the given forces, the tongue had some extreme deformations, but the deformation was smooth and no divergence was seen. In these evaluations, the proposed model demonstrated excellent performance on both static and dynamic behaviors.

D. Collision of the tongue and the vocal-tract wall

In speech articulation, the tongue often contacts the teeth, hard palate, and jaw as it moves. The tongue tip, for example, collides with the vocal-tract wall when producing consonants such as /t/ and /l/. The lateral parts of the tongue contact the hard palate to form a narrow airway of the vocal tract in producing the vowel /i/ or alveolar fricatives. Thus, the contact of the tongue with the vocal-tract wall is one of the important factors in achieving accuracy and stability in the dynamic control of the tongue. It also generates external forces that affect tongue deformation when the contact occurs. Therefore, the realization of tongue–wall contact is an essential task for a physiological articulatory model.

Since the shape of the vocal-tract wall is too complex to be described by an analytic function, the contact of the tongue with the vocal-tract wall, unlike the other constraints, cannot be combined into the motion equations systematically. As an alternative, we propose a method with three steps to deal with the contact between the tongue and the tract wall to check whether or not the nodes of the tongue cross through the tract wall; find the equilibrium position on the wall for the nodes outside of the vocal tract, and distribute the collision force. If a node crosses through the wall of the vocal tract during articulation, its trajectory must have an intersection with the tract wall. Since the tract wall was assembled by triangle planes, we first identify the plane with which the trajectory intersected on the wall and then calculate the collision force of the node when it is bounded on the wall. The following formula is used to estimate the collision force:

$$f_x = \sum_i (k_i \Delta l_{xi} + b_i \Delta l_{xi}/h), \tag{10}$$

where i is the index of the cylinders connected to the node, and h is the computation step. k_i and b_i are the stiffness and viscous coefficients of cylinder i. Δl_{xi} is the increment of cylinder i in the x dimension caused by the wall bounding. f_x is the x component of the resultant bounded force. Using the same approach, the force can be calculated for the y-dimension f_y and z-dimension f_z. To reach an equilibrium position on the wall, the coordinates (x, y, and z) of the node must meet the following simultaneous equations:

$$\begin{aligned} &A(x-p_x)+B(y-p_y)+C(z-p_z)=0,\\ &\alpha(x-p_x)=f_x,\\ &\alpha(y-p_y)=f_y,\\ &\alpha(z-p_z)=f_z, \end{aligned} \tag{11}$$

where p_x, p_y, and p_z are the coordinates of the intersection on the plane; A, B, and C are the norm of the plane; and α is an unknown equivalent factor of the stiffness and viscosity. Since such a deformation cannot be predicted in the motion equations, the collision forces must be considered additionally. In model calculation, the collision force above is taken into account at the next computation step as an input.

When the tongue slides on the wall surface to reach an equilibrium position, friction between the tongue and the palate is a considerable factor. This friction can be approximated as the force generated in laminar flows of the mucosa, since there is much mucosa on the surfaces of the tongue and the hard palate. Such a friction force is proportional to the viscosity of mucosa, the contact area between the tongue and palate, and the velocity of the tongue, but is inversely proportional to the thickness of the mucosa. To obtain the order of the magnitude for such friction, we roughly estimate the friction force based on the following conditions. Suppose the mucosa has about the same viscous coefficient as water, which is 0.01 dyne·s/cm^2 at 20 °C, and the thickness of the mucosa between the tongue and palate is very thin, about 0.01 cm. The maximum velocity of the tongue is about 40 cm/s (see Table II). Under the given conditions, the force caused by the friction for a unit area is about 40 dynes, or 4×10^{-4} N. This friction force is much smaller than the forces caused by muscle contraction and wall reaction. Therefore, this friction is ignored at the current model. Since the friction can be expected to stabilize the tight constriction to some extent, it may be helpful in achieving accurate control.

IV. ESTIMATION OF MUSCLE FUNCTION BY MODEL SIMULATIONS

As stated in the Introduction, many studies have sought to ascertain the relationship between tongue movement and muscle activation. However, it is generally difficult to determine such a relationship from experimental observation alone because critical parameters such as mechanical load on a muscle cannot be observed. This section examines muscle functions using model simulations.

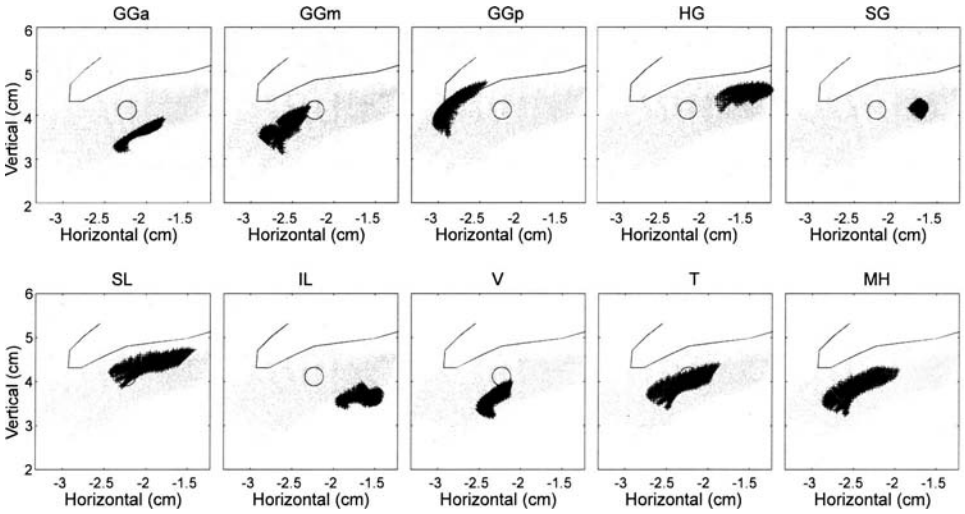

FIG. 5. Starting point region of the tongue tip, shown by the pale small circles, and the ending points driven by a force of 4 N with a 150-ms duration, shown by the dark crosses.

A. Muscle activations and tongue movements

In order to evaluate tongue muscle functions by simulation, tongue movements are represented using two points of the tongue tip and the tongue dorsum, which are referred to as the *control points* (Dang and Honda, 2001; 2002).

1. Function of tongue muscles

In producing an utterance, the tongue can start its movement from many different positions or shapes. The coming deformation of the tongue involves all of the past deformations and/or the history of the muscle forces to some extent. To account for the effects of the past deformations on tongue behaviors, a scattered region of starting points was designed for the control points, based on observations obtained from the prototype speaker of the model using the x-ray microbeam system (Hashi *et al.*, 1998). The control points were first moved from the initial position to a given starting point. The initialization movements were governed by four muscle combinations: GGp–GGm–GGa, GGp–SG–GGa, HG–GGm–GGa, and HG–SG–GGa. Activation forces for each muscle were set to be six levels between 0 and 6.0 newtons with a 120-ms duration. Nonlinear intervals were adopted in the six levels to achieve a uniform distribution. As a result, 864 starting points were generated for each control point, as shown by the small gray circles in Figs. 5 and 6 as a reference.

After the control points arrived at the given starting points, all forces were released and then the tongue body was driven by a specific muscle with a given force. The force was 4 N with a 150-ms duration. Figure 5 plots the ending locations of the tongue tip, shown by the dark crosses. The small gray circles show the starting points, and the large circle denotes the initial location of the control point. Note that the boundary and contact force of the tract wall were not considered in estimating the muscle functions.

As shown in this figure, even though the tongue tip started from widely scattered points, it converges to a specific, small region for all of the muscles. When the GGa contracts, the tongue tip concentrates to a strip of a region that is lower than the initial position, shown by the large circle. This means that the function of the GGa is to lower the tongue tip. Similarly, the GGm moves the tongue tip forward and downward, while the GGp drives the tongue tip forward and upward. The HG moves the tongue tip backward and slightly upward. The SG mainly drives the tongue tip backward, where the convergence region for the SG is smaller than that of the others. This implies that the SG has

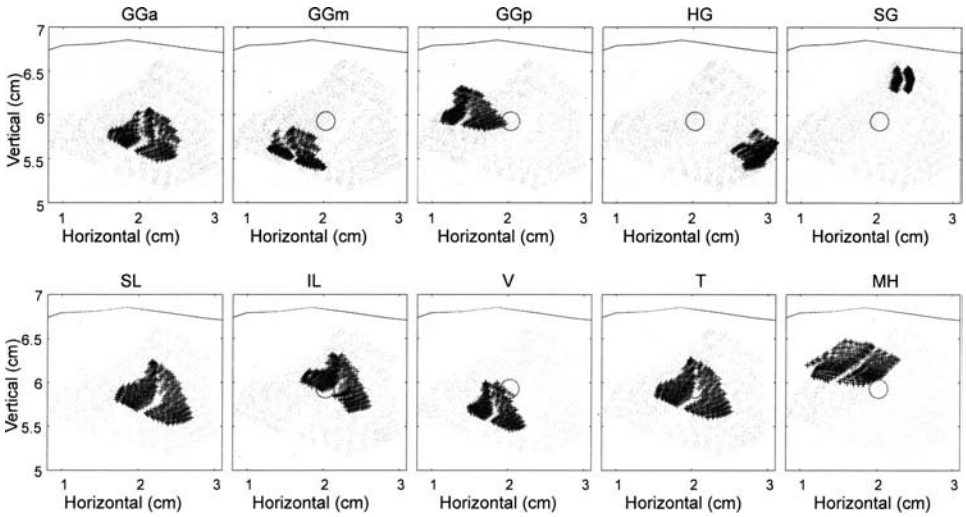

FIG. 6. Starting point region of the tongue dorsum, shown by the small gray circles, and the ending points driven by a force of 4 N with a 150-ms duration, shown by the dark crosses.

a clearer equilibrium position for the tongue tip, or that this equilibrium position is reached faster than the others.

For the intrinsic muscles, the inferior longitudinalis (IL) has a definite function that moves the tongue tip backward and downward, while the superior longitudinalis (SL) drives the tongue tip upward and backward. The verticalis (V) moves the tongue tip downward. The transversus (T) and mylohyoid (MH) did not show any definite contribution to movement of the tongue tip. The geniohyoid is not shown in the figure because its contribution to both the tongue tip and the dorsum was small, or underestimated in the model simulation.

Figure 6 shows the scattering of the starting points and ending locations for the tongue dorsum, where the simulation condition was the same as that for Fig. 5. Among the extrinsic muscles, the GGm, GGp, HG, and SG show definite functions that drive the dorsum to go front-lower, front-upper, back-lower, and back-upper, respectively. These results are consistent with the findings in previous studies (cf. Baer *et al.* 1988), while the GGa did not show any significant effect on the dorsum movement. The MH has a certain effect on the tongue body, which drives the tongue dorsum upward and forward. The verticalis moves the dorsum downward and forward to a certain extent, while the IL moves the dorsum backward and upward. The SL moves the dorsum slightly backward and downward, while the transversus did not show any contribution to the dorsum movement.

Putting Figs. 5 and 6 together, one can see that when the HG is activated the tongue dorsum moves backward and downward, while the tongue tip moves backward and slightly upward. This suggests that an accompanying rotation of the tongue body takes place during such movements. The above simulation shows that when the same muscle forces are applied on the tongue, the control point from different starting positions converges to a region but not a point during the given duration. Since the duration is close to vowel duration in speech with a normal rate, this suggests that the observed vowel target regions in the electromagnetic articulographic data (cf. Dang *et al.*, 2002) may be a consequence of tongue biomechanics. The simulation results also suggest that the tongue muscles may need a longer time to achieve their equilibrium position.

2. Muscle forces and equilibrium position

The results plotted in Figs. 5 and 6 show that the control points were driven to arrive at specific regions in spite of their widespread starting points. For the purpose of model control, we consider two issues. One is whether the control points equilibrate at one point or converge to a sufficiently small region when the activation duration is sufficiently prolonged. The other is whether the relationship between a muscle force and an equilibrium position is unique.

A numerical experiment was conducted to answer these two questions. Four starting points were generated by four muscle combinations with a force of 5.0 N for all muscles. The tongue was then driven from the starting points by a specific muscle with a 400-ms duration. The activation forces were 1.0, 2.0, 3.5, and 5.0 N. This simulation was carried out on all of the tongue muscles. Figure 7 shows

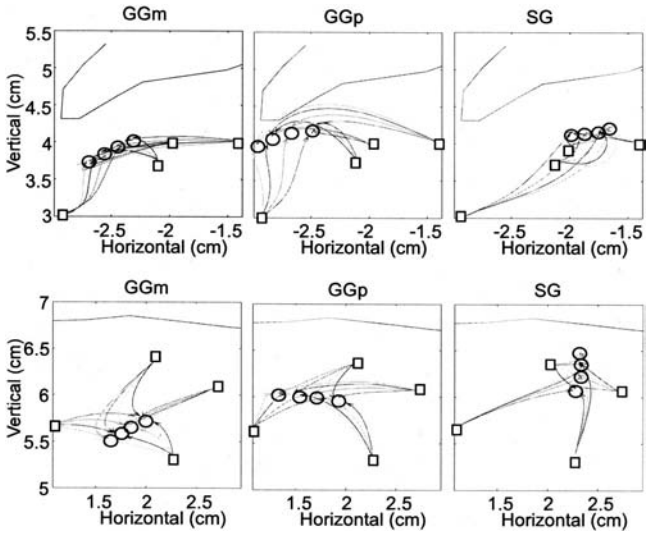

FIG. 7. Trajectories of the tongue tip (upper) and the dorsum (lower) from distinct starting points (indicated by squares) to certain ultimate locations (shown by circles) that correspond to the force levels of 1, 2, 3.5, and 5 N.

examples for three extrinsic muscles. The upper panels demonstrate the results for the tongue tip, and the lower panels are for the tongue dorsum. For each starting point, the trajectories of the control points spread out along different paths corresponding to the force levels. For different muscles, the trajectories have dissimilar curvatures. When activating the GGp, for example, both the tongue tip and the dorsum tended to reach force-dependent locations in a straight path. In the cases of the GGm on the dorsum and the SG on the tongue tip, the control points move along curved paths from the starting points to the final locations. The difference between the paths is mainly caused by previous deformation (starting point) and/or the history of the forces. However, they finally converge at one point, where the model reaches an equilibrium state. In the simulation, the control points generally reach their equilibrium positions within about 300 ms. These final points are referred to as the *equilibrium positions* hereafter. It is interesting to find that the equilibrium position of each muscle shifts monotonically as the force level increases. This means that the equilibrium position and the force level have a unique relation for a given muscle.

B. Mapping between equilibrium position and articulatory target

In order to control the model via the muscle forces, it is useful to find a mapping between the muscle forces and articulatory targets. Such a mapping is constructed in this section based on the equilibrium positions.

1. Equilibrium positions of muscles

As shown in Fig. 7, the equilibrium position (EP) for each muscle corresponds to its activation level, despite past deformations. This relation provides a connection between a muscle force and a spatial point in the articulatory space which is invariant for a given muscle structure. Using such a connection, a unique mapping can be obtained from a muscle force to a spatial position. However, the inverse mapping

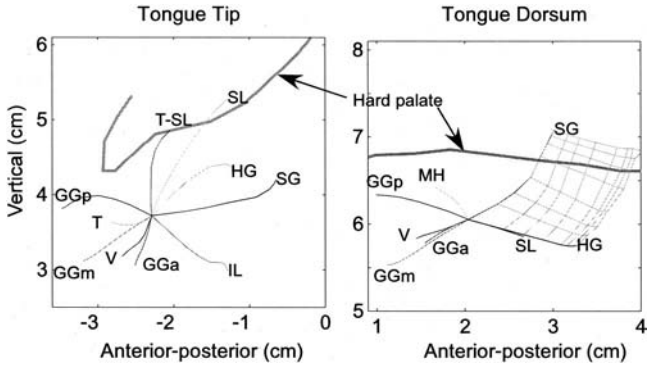

FIG. 8. Coordinates consisting of the equilibrium positions corresponding to the activation forces ranged between 0 and 6 N. The net in the right panel consists of the contour lines of the EPs of SG and HG.

TABLE III. Contribution factors and the major contributions of the tongue muscles.

Muscle names	Tongue tip	Tongue dorsum	Contribut. (>0.3)
Genioglossus A.	0.37	0.39	Both
Genioglossus M.	0.58	0.72	Both[a]
Genioglossus P.	0.68	0.74	Both[a]
Hyoglossus	0.74	0.86	Both[a]
Styloglossus	0.94	1	Both[a]
Superior longitudinalis (SL)	1	0.45	Tip[a]
Inferio longitudinalis	0.69	0.13	Tip[a]
Verticalis	0.36	0.43	Both
Transversus (T)	0.31	0.22	None
T-SL group	0.65	0.17	Tip[a]
Myohyoid	0.21	0.36	Dorsum

[a]Shows the muscles whose contribution factor is larger than 0.5.

from spatial position to muscle force is not unique, which brings about the ''one-to-many'' problem. Note that the *equilibrium position* depends on the muscle activation level alone, where no hypothesis is involved with the threshold of the stretch reflex. To distinguish the proposed method from the *equilibrium point hypothesis* (EPH, λ model), we use the term of *equilibrium position* hereafter, but not *equilibrium point*.

In order to obtain a series of EPs, each muscle is activated by an altering force for 300 ms, where the force was applied at levels of 0.0, 0.1, 0.2, 0.4, 1.0, 2.5, 4.0, and 6.0 N, respectively. Thus, a coordinate is established based on the EPs for each control point. Figure 8 shows the coordinates for the tongue tip and tongue dorsum. Since the EPs shift monotonically, the equilibrium position for a muscle can be expected to move along the path consisting of the EPs as the muscle force varies continuously, as long as the other muscles' forces remain unchanged. The path built on the EPs can be considered a ''vector'' that spreads out from the rest position. The HG moves the tongue tip backward and slightly upward as the activation level increases. The SG drives the tongue tip backward almost horizontally, while the GGa draws the tongue tip downward. The tongue tip is driven forward-upward by the GGp and forward-downward by the GGm. The intrinsic muscles SL and IL move the tongue tip largely upward-backward and downward-backward, respectively.

In Fig. 8, one can see that if the T-SL muscle group is not considered, there is a large blank space between the vectors of the SL and GGp in the tongue tip coordinate. This means that no single muscle can move the tongue tip in that direction. Since this region is important for constructing the alveolar consonants, we have proposed a muscle group consisting of the transversus (T) and the SL to fill this blank space with a new EP vector of T-SL. After this muscle group is added, the EP vectors are distributed almost uniformly in the coordinate for the tongue tip. Such a coordinate makes it possible to move the tongue tip involved in any direction.

The right panel of Fig. 8 shows the EP vectors for the dorsum. Unlike the tongue tip, the dorsal coordinate has a much larger scale in the horizontal direction than in the vertical direction. The extrinsic muscles (except the GGa) have definitely larger EP vectors than the others. The EP vectors of the GGp and HG show diametrically opposing directions. This indicates that the GGp and HG work antagonistically in governing the tongue dorsum. Similarly, the GGm and SG form another antagonist muscle pair. Therefore, there are obviously two large antagonist muscle pairs controlling the dorsum.

As the dorsum is driven upward and backward by the SG, it begins to cross through the hard palate when the muscle force reaches 1.0 N. The simulation shows that the model can make the gesture for /k/ with a force of about 1.0 N when the jaw is in the rest location. For the GGp, its EP vector did not reach the hard palate in the dorsal coordinate, while a contact between the tongue dorsum (or the blade) and the anterior portion of the hard palate was required in a number of articulations. As shown in Fig. 4, the tongue dorsum actually had reached the hard palate within 20 ms in the simulation, but the control point (the fifth node from the tongue tip) could not reach the palate. This is because the highest point of the dorsum did not always correspond to a fixed node of the model during the deformation, as demonstrated in the left and right panels of Fig. 4. This study uses the fifth node to represent the dorsum since the highest point is around the node in the most cases.

2. Construction of the EP map

In the proposed model, there are 11 muscles involved in the tongue body. To simplify descriptions, this study focuses on the major muscles only. For this purpose, muscle contributions are evaluated using the amplitude of the EP vectors, that is, the shift distance of the EP as the force increases from 0 to 6 N.

Among the EP vectors, the SL has the largest amplitude in the tongue tip coordinate, while the SG generates the largest vector for the tongue dorsum, as shown in Fig. 8. The vector amplitudes are normalized by the maximum one for the tongue tip and dorsum, respectively. The normalized amplitudes reflect a relative contribution of the muscles to tongue movements, and they are referred to as the *contribution factor*. Table III shows the contribution factor of the muscles. A muscle is considered to have a certain contribution to the tongue movement if its contribution factor is larger than 0.3. The superscripted [a] denotes the muscles with a contribution factor larger than 0.5. As a result, the most

extrinsic muscles (except the GGa) have larger power over both the tongue tip and tongue dorsum. Among the intrinsic muscles, the SL and IL show definite contributions to the tongue tip. The transversus demonstrates a certain contribution when it is grouped with the SL, which plays an important role in increasing the control freedom. The geniohyoid showed no significant contribution to any control point. Note that this study only used two specific points to evaluate the contribution of the muscles. If different observation points were adopted, a dissimilar contribution factor might be obtained for some muscles, e.g., the geniohyoid.

To develop a control method, it is necessary to reinspect the EP vectors based on the contribution and the freedom of model control. All of the EP vectors shown in Fig. 8 have a contribution factor larger than 0.3. In the coordinates of the tongue tip, the GGa and verticalis (V) have a similar vector with smaller amplitude. These two vectors are taken into account because they increase the degree of the freedom. The transversus (T) is also taken into account because it increases the degree of freedom, although its role is not explicit in a 2D representation.

In the dorsal coordinate, the EP vector of the MH is located in the space between GGp and SG and contributes to widening the area of contact between the dorsum and the palate, which is required for generating /k/ gestures with different contexts. For this reason, the MH is included in the dorsal coordinate for the control, although its vector is smaller than that of the extrinsic muscles. The GGa and V have a similar EP vector to that of the GGm, but their amplitude is about half that of the GGm. These muscles are not considered an independent factor in the control method because they neither contribute significantly nor increase the degree of freedom. For the same reason, the SL is not treated as an independent factor. However, these muscles are taken into account in the muscle co-contraction, described in the subsequent section.

Based on the above consideration, five major muscles are taken into account in controlling the tongue dorsum, and nine muscles and one muscle group are used for the tongue tip. Thus, the mapping between the spatial points and the muscle forces can be obtained based on the selected EP vectors. An example is shown in the right panel of Fig. 8 by a contour net, which consists of the EPs of the SG and HG. The contour lines correspond to the six force levels. Such a net of contour lines is named the *equilibrium position map* (EP map). With the EP map, any arbitrary point inside the region of the map can be reached using the forces interpolated from the contour lines. The primary difference between the EP map and the EPH (λ model) is that the EP map is a straightforward mapping between muscle forces and the equilibrium positions of the articulators, while the EPH not only requires a muscle loading function but also the λ commands that are involved in a hypothesis on the threshold of the stretch reflex.

3. EP map of the tongue vs jaw positions

Since there is a high correlation between the movements of the tongue and jaw, the coordinate system consisting of the EPs depends strongly on the jaw position. To examine the relationship between the EP maps and jaw positions, we first investigated the equilibrium position of the jaw complex. Five forces of 0.5, 1.2, 2.5, 4.0, and 6.0 N with a duration of 300 ms were applied to the jaw-opener muscle individually, and three forces of 0.5, 1.5, and 3.0 N were applied to the jaw-closer muscle. Corresponding to the forces on the jaw opener, the jaw achieved EPs of -0.16, -0.37, -0.76, -1.10, and -1.42 cm, where the minus sign denotes a position lower than the rest position. When the jaw closer was activated, the jaw reached EPs of 0.09, 0.21, and 0.47 cm in the close direction. Altogether, nine equilibrium positions were obtained for the jaw, which includes one rest position, five opener positions, and three closer positions. The EP maps for the tongue were constructed corresponding to the nine positions of the jaw.

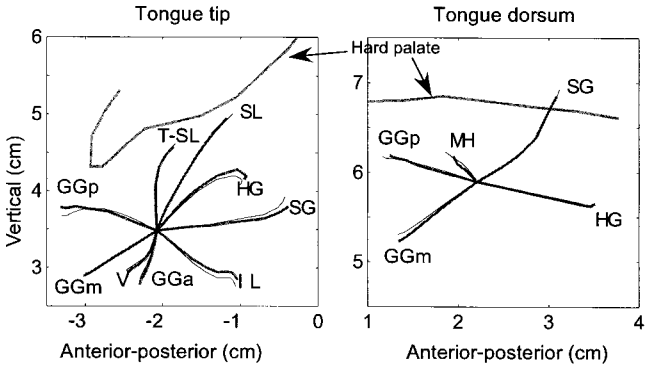

FIG. 9. Example of using a dynamic EP map (the thin lines) derived from the typical EP map of the jaw opener at 0.5 N to represent the EP map (the thick lines) of the jaw opener at 2.5 N.

Comparison of the nine jaw-dependent EP maps shows that the entire EP map rotates and translates with the jaw movement, but the detailed structures of the EP maps has no definite changes. This suggests that it is possible to derive a *dynamic* EP map for any arbitrary jaw position from one EP map by means of translation and rotation. To obtain an optimal dynamic EP map for all the jaw positions, the EP map generated by applying 0.5 N on the jaw-opener muscle is chosen as the typical one, since it is the neutral one in the nine positions. Figure 9 shows an example using a dynamic EP map (the thin lines) derived from the typical EP map to represent the EP map (the thick lines) of the jaw opener with 2.5 N. The rotation degree for the dorsal coordinate was the same as that of the jaw, while a slight inverse rotation was carried out for the tongue tip. In this representation, the average difference between the two EP maps was 0.049 cm for the tongue tip and 0.041 cm for the dorsum.

The accuracy of representing the jaw-dependent EP maps by the typical EP map was evaluated by a mean-squared error between the derived dynamic EP maps and the original ones. Table IV shows the evaluation for all cases, where the top corresponds to the closed positions of the jaw, and the bottom to the wide-open jaw position. The average error over all jaw positions is 0.043 cm for the tongue tip and 0.034 cm for the dorsum. The error increases as the jaw deviates from the reference position. The largest error of 0.12 cm occurred for the tongue tip when the jaw opened about 1.8 cm. For almost all articulations, the tongue tip is not so

TABLE IV. Differences between derived and original EP maps.

Status of the jaw	Force on the jaw muscle (N)	Position on the initial (cm)	Errors for the apex (cm)	Errors for the dorsum (cm)
Closer	3.0	0.407	0.037	0.030
Closer	1.5	0.210	0.026	0.024
Closer	0.5	0.090	0.021	0.020
Initial	0.0	0.00	0.024	0.019
Opener[a]	0.5	−0.160[b]	0.000	0.000
Opener	1.2	−0.374	0.021	0.018
Opener	2.5	−0.761	0.049	0.041
Opener	4.0	−1.097	0.084	0.066
Opener	6.0	−1.425	0.123	0.091

[a]Indicates the typical EP map.
[b]The minus denotes jaw positions lower than the rest position.

crucial when the jaw opens wide. In other words, this representation does not introduce any significant error for crucial points. Figure 9 and Table IV illustrate that a dynamic EP map can be obtained effectively for any jaw position via this derivation.

V. CO-CONTRACTION OF MUSCLES BASED ON THE EP MAP

The most effective way to form a tongue shape by muscle contraction, in view of the minimal energy principle, is for two agonist muscles to work together to achieve a given target. Thus, the muscle forces can be easily estimated via the EP map shown in Fig. 8. During speech, however, the situation is much more complicated because more than two agonist muscles can work together to reach a target, and some agonist–antagonist muscles can co-contract at the same time. To simulate this situation, we designed 17 two-muscle groups and seven three-muscle groups for the tongue tip and five two-muscle groups and one three-muscle group for the dorsum. An example of the EP map for the two-muscle group is shown in the right panel of Fig. 8. The three-muscle group consists of an independent muscle and a muscle pair, in which the activation of the independent muscle corresponds to the co-contraction level and governs a *main* part of the tongue, while the muscle pair manipulates the other part via the mechanism of co-contraction of the agonist and antagonist muscles.

A. Co-contraction between agonist and antagonist

Eight three-muscle groups were designed to generate some potential co-contractions during speech. Figure 10 shows the co-contractions for two three-muscle groups. The thick dark lines show a part of the EP vectors of the coordinates, which were generated by activating the muscles individually. The thin dark lines and the thin gray lines denote the EP trajectories for two synergistic muscles in the group, respectively. The attachment of the thin lines on the EP vector of the independent muscle, indicated by the open circles, corresponds to activation levels of the independent muscle. As shown in the upper panels, the combination of the SG and the muscle group of the GGp and SL can move the dorsum toward the palatal target by the SG and at the same time control the tongue tip to an apical target by the muscle pair. This mechanism can reach a compatible target set for both

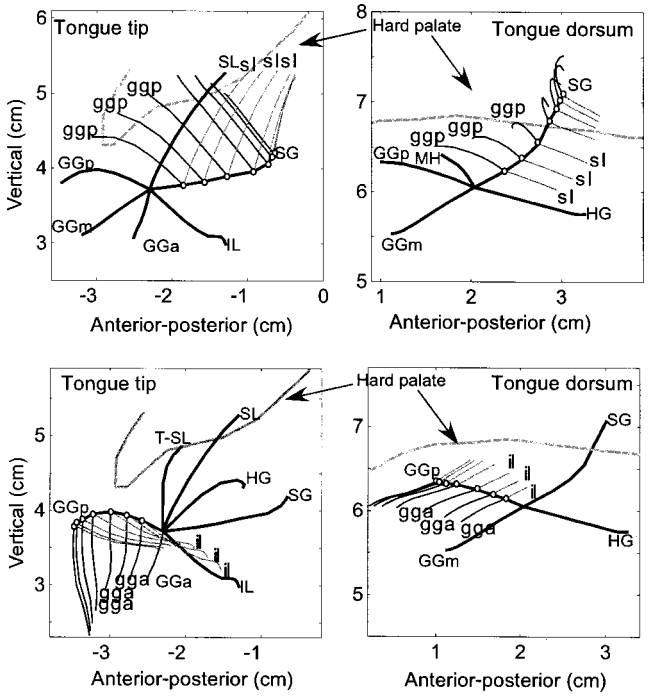

FIG. 10. Co-contractions of the SG and a muscle pair of the GGp and SL (upper panels), and of the GGp and a muscle pair of the GGa and IL (lower panels).

the tongue tip and the dorsum. It is interesting to find that the muscle pair works in synergy for the tongue tip while it functions as an antagonist pair for the tongue dorsum. If the proper force ratio is chosen for the muscle pair of the GGp and SL, the dorsum can be kept in a given position when the tongue tip position is manipulated, and vice versa.

The lower panels in Fig. 10 show the co-contraction between the GGp and the muscle pair of the GGa and IL. This muscle group drives the dorsum to move in a front-upper direction and governs the tongue tip to go downward and backward. The muscle pair demonstrates the same function as that in the upper panel, in which the muscle pair shows a synergetic function for the tongue tip while it behaves as an antagonist pair for the tongue dorsum. Using this combination, the tongue tip can be retracted without interference to the dorsum because the GGa and IL work as antagonists for the dorsum.

Figure 11 shows other examples of the co-contractions. In the upper panels, the combination of the GGm and a muscle pair of the SL and T-SL demonstrates a function that produces a posture with a lower dorsum and a higher tongue tip. The tongue tip is governed by the muscle pair of two intrinsic muscles, SL and T-SL, and an optimal dorsum position can be maintained by using a proper force ratio for the muscle pair. Compared to the other cases, the intrinsic muscles have a smaller effect on the dorsum. It can be seen that the intrinsic muscle pair also definitely works as an antagonist pair for the dorsum. The lower panels of the figure show a different combination, the T-SL and a muscle pair of the GGm and HG, in which the independent muscle is an intrinsic muscle, and the muscle pair consists of two extrinsic muscles. In contrast to the other groups, this group first governs the tongue tip by the intrinsic muscle and then ma-

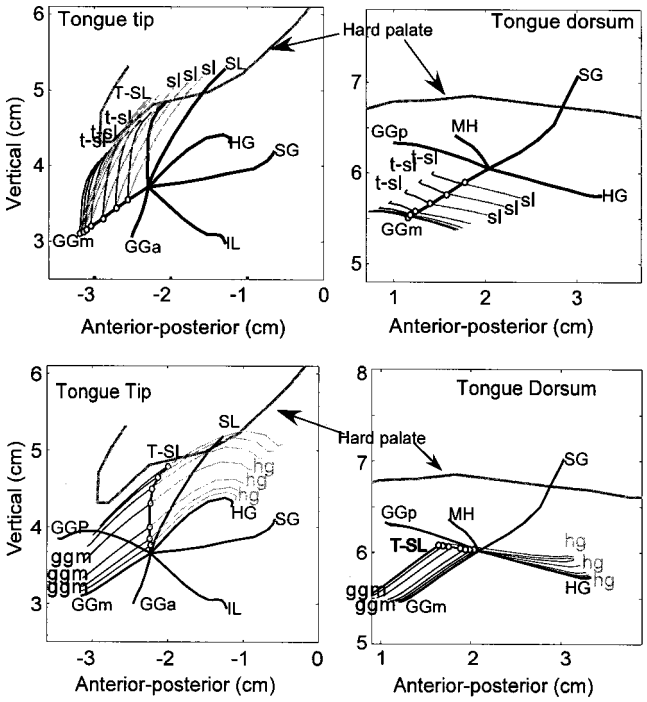

FIG. 11. Co-contractions of the GGm and a muscle pair of the SL and T-SL (upper panels), and the T-SL and a muscle pair of the GGm and HG (lower panels).

nipulates the dorsum by the extrinsic muscles, where the two extrinsic muscles function as an antagonist muscle pair for the tongue tip.

In all of the three-muscle groups, the muscle pair works in synergy for one part of the tongue while behaving as an antagonist pair for the other part. This function can increase the degree of the freedom for model control, allowing us to control different parts of the tongue independently to some extent. Because of this property, we might say that the co-contraction can also be used to maintain the stability of a kinematic system when part of the system is manipulated.

B. Realization of multiple features using co-contraction mechanism

Generally, in a target vector, only one crucial feature is decisive in forming the phoneme, where the other features are referred to as *indecisive* features. To produce a planned sound, the accuracy of the crucial feature must be guaranteed, while an optimal achievement is also desired for the indecisive features. This requires us to deliberate over more than one feature at the same time to find a force set that is optimal for the concerned target features. Figure 12 shows an example of realizing this procedure using a three-muscle combination consisting of the HG and a muscle pair of GGp and SL. In this example, we estimate muscle forces for a given consonantal target with two features for the tongue tip and dorsum, respectively, where the feature for the jaw position was taken into account in generating the jaw-dependent EP map. The crucial feature for the tongue tip is shown by the filled circle, and an indecisive feature for the dorsum is indicated by the larger open circle, which accounts for the coarticulatory effects of the surrounding vowels.

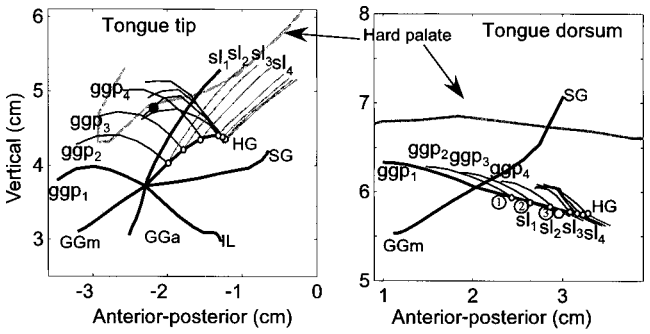

FIG. 12. Setting the apical and dorsal targets simultaneously by means of the co-contraction of the HG and a muscle pair of the GGp and SL. The circles denote the targets for the control points. The circled numbers indicate the candidate targets corresponding to different co-contraction levels.

In Fig. 12, one can see that the HG is the major muscle driving the tongue dorsum toward the dorsal target, while the muscle pair of the GGp and SL is one of the muscle groups that can drive the tongue tip to reach the apical target. Among the force combinations, for example, three combinations of the GGP_1-SL_1, GGP_2-SL_2 and GGP_3-SL_3 are capable of moving the tongue tip to reach the apical target. The difference between them is that they correspond to different co-contraction levels of the HG, whose forces were 0.0, 0.1, and 0.2 N. Since all three combinations can guarantee the crucial feature, the decision of the force set finally depends on their behavior on the indecisive feature. The circled numbers in the dorsal coordinate are the predicted locations for these three force sets. The location of ③ is the closest one to the given dorsal target among the three sets. Therefore, the force set of the GGP_3-SL_3 and HG with 0.2 N is the optimal one for the given target with two features.

In a general process for estimating activation patterns, all possible force combinations are searched out through the EP maps for a given target. The square summation of the muscle forces is calculated for each combination, and the distance between the given target and the candidate target is computed for the indecisive features. The cost function for determining the force set is the weighted summation of the square sum of the forces and the distance. The principle for determining the force set is that it should accurately guarantee the primary feature and optimally realize the indecisive features.

VI. SUMMARY AND DISCUSSION

This study consists of three parts: (1) improve the modeling of tongue tissue; (2) investigate the muscle functions using model simulation; and (3) develop an estimation method for muscle activation forces from spatial targets while considering the co-contraction of muscles.

A. Improvement of the model

The previous version of our model (Dang and Honda 2001; 2002) used volumeless springs as a mesh component, which are replaced by viscoelastic cylinders in the current version. The cylinder is a continuum that fully obeys the physical laws described by the Young's modulus and the Poisson ratio. After dividing the tongue tissue into a hexahe-

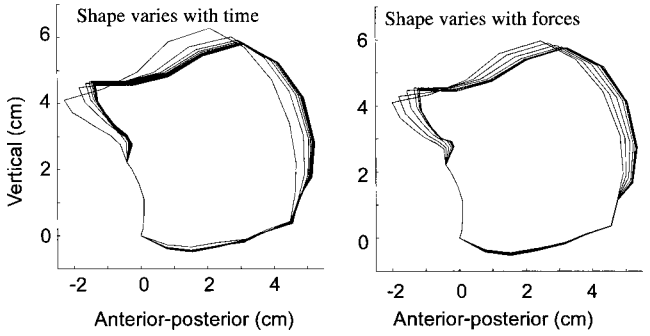

FIG. 13. Temporal deformation of the tongue when a force of 1.5 N is applied on HG (left panel), and the tongue shapes when the HG reaches the equilibrium position for given forces from 0.1 to 4.0 N (right panel).

dron assemblage, each hexahedron was filled with 28 cylinders connecting its vertices in all possible directions. This assemblage of cylinders provides an efficient way to describe the deformations of a hexahedron, because the force applied on any node is decomposed and shared by all connecting cylinders. The continuum properties were achieved to some extent by using the Poisson ratio in the volume constraints of Eq. (6). However, the truss of the cylinders is not a continuum but is discrete. This may introduce some differences from the standard FEM. The result obtained from the test cuboid showed that the semicontinuum model is sufficient for modeling the continuum body for a first-order approximation.

The major benefit of using the semicontinuum model is that a fast computation can be achieved without any significant loss in accuracy. Computation times were about 40 times the real time for simulating articulatory movements, which are more than two orders of magnitude faster than that using the standard FEM. Another advantage is that the truss–frame structure is appropriate for large deformations such as tongue movements. No divergences occurred even when a very large force was applied on the proposed model or an extreme deformation took place during articulatory movements.

B. Behavior of the muscles

This study demonstrated a systematic simulation for all of the tongue muscles. The muscles' contribution to tongue movements was evaluated by the normalized amplitude of the EP vectors. Most extrinsic muscles (except the GGa) govern both the tongue tip and tongue dorsum. The SL and IL showed clear contributions to the tongue tip movements.

The simulation revealed some differences between the results of the present study and those of other studies. First, the HG moves the tongue tip backward and slightly upward but not downward. This is somewhat different from the commonly held view that the HG may also lower the tongue tip. Figure 13 shows some simulations to clarify the causes. The left panel demonstrates the temporal deformation of the tongue when a force of 1.5 N is applied on the HG, and the right panel shows the tongue shapes when the HG reaches an equilibrium position for given forces from 0.1 to 4.0 N. The simulation shows that the tongue moves backward and downward with a slight rotational motion under both conditions. This rotation might explain why the HG does not produce a downward movement of the tongue tip. In tagged MRI observations, Davis *et al.* (1996) found that a local depression occurred behind the tongue tip for steady-state /a/ and suggested that it is due to a shear deformation of the HG with GGa contraction. Analysis of the tagged cine-MRI supported the idea that the GGa produces a tongue-lowering motion during the production of /a/ (Stone *et al.*, 2001). It seems that when the tongue tip moves downward with the tongue dorsum the GGa contracts, assisting the HG.

For the function of the GGp, previous observations showed that the GGp moves the tongue dorsum first forward and then upward (Baer *et al.*, 1988; Alfonso *et al.*, 1987). Our model simulation could not confirm this phenomenon. As shown in Fig. 7, the temporal trajectory did not show any definite tendency in the temporal order of the upward and forward movements resulting from the GGp. The simulation showed that the trajectories depend almost entirely on the force level and the history of the deformation.

A few previous studies stated that the SL, in combination with the GGa, could elevate the tongue tip (Stone *et al.*, 2001; Napadow *et al.*, 1999). This was based on two observations: GGa fibers do not curve forward or extend into the tongue tip, and GGa contraction locally stiffens the tongue behind the tongue tip. Our simulation showed that the SL moves the tongue tip upward, while the GGa has an antagonistic function with the SL to move the tongue tip as shown in Fig. 8. From the point of view of the minimum energy principle or co-contraction, it seems difficult to find a reason for the use of these two muscles as a pair to elevate the tongue tip. However, we cannot negate the suggestion by other studies since the tongue blade was not observed in the current study.

C. Mapping from spatial targets to muscle forces

The equilibrium position for each muscle monotonically shifts as the muscle force increases. Based on the equilibrium position, the EP map was proposed as a way to construct a straightforward mapping between the spatial positions and muscle forces. The force estimation based on the EP map focuses on the spatial target, i.e., the ending point, alone. The estimated force activates a muscle constantly until a new target is coming. Such a force actually is a static force, since dynamic movements of the model were not considered during either force estimation processing or the activation period. For certain large deformations, as shown in Fig. 7, the trajectory is not a straight line. It is difficult to define such a curved trajectory by using its starting and ending points only. Actually, the trajectory can also be another measure for the spatial target (Okadome and Honda, 2001). Achieving trajectory control also requires us to estimate the dynamic component of the forces. The authors have developed a method, namely *muscle workspace*, to estimate the dynamic muscle forces stepwise by minimizing the distance from a current position and the target (Dang and Honda, 2002). Combination of both of the static and dynamic force estimations in the model control method remains a development task for future work.

To simulate co-contractions between agonist and antagonist muscles, we designed several three-muscle combinations, consisting of one independent muscle and a pair of two muscles. The simulation showed that the two muscles of the muscle pair worked in synergy for one control point while functioning as an antagonist muscle pair for the other one. This function can be used to keep the stability of a kinematic system when a part of the system is manipulated. This property can be used on the model to achieve multiple targets simultaneously, based on a strategy that accurately guarantees the crucial target and optimally reaches the indecisive targets.

A method was proposed in this study to estimate the muscle patterns for a given target based on the EP maps, in which all possible force combinations are searched out through the EP maps, and then one optimal force set is chosen to accurately guarantee the primary feature and optimally reach the indecisive features. Model simulation showed that this method is an efficient way to achieve multiple targets by using the co-contraction mechanism. However, our proposed method may be far different from that used by the human brain, which is still too complicated for humans to learn and use to optimize muscle activation patterns. Therefore, a more efficient method is still desired.

ACKNOWLEDGMENTS

A part of this study was finished when the first author was working for the Institut de la Communication Parlee (ICP) Grenoble, France. The authors would like to thank Pascal Perrier for general discussions as well as thank Mark Tiede, Donna Erickson, and Pierre Badin for their valuable comments. The authors deeply appreciate their instructive comments of Anders Löfqvist and two anonymous reviewers of this paper. This research has been supported in part by the Telecommunications Advancement Organization of Japan.

APPENDIX: ANALYSIS OF DISPLACEMENT-BASED FINITE ELEMENT METHOD

The present study adopts a displacement-based finite-element method as the basis of the modeling, which is referred to as an extended finite element method (X-FEM). The principal advantages of X-FEM are that the finite-element framework (sparsity and symmetry of the stiffness matrix) is retained and that a single-field variational principle is used (Zienkiewicz and Taylor, 1989; Belytschko *et al.*, 2002).

To apply the X-FEM to our modeling, the equilibrium of a general 3D body is first considered, as shown in Fig. 14. The body is located in the fixed coordinate system *X, Y, Z*. The body surface area is supported on the area $\mathbf{S}_u$ with prescribed displacements $\mathbf{U}^{Su}$ and is subjected to surface traction $\mathbf{f}^{Sf}$ on the surface area $\mathbf{S}_f$. In addition, the body is subjected to the externally applied body force $\mathbf{f}^B$ and concentrated loads R^i (where i denotes the point of load application). The displacements of the body from the unloaded configuration are measured in the equation

$$P(X,Y,Z)=[U,V,W]^T, \tag{A1}$$

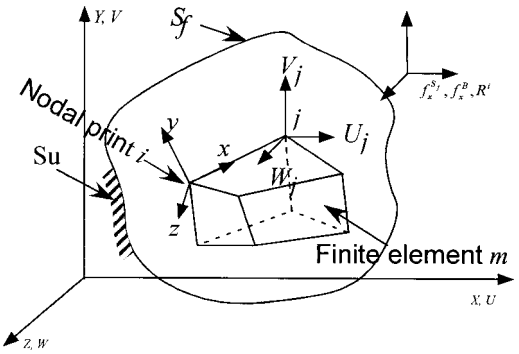

FIG. 14. General three-dimensional description of the displacement-based finite-element method.

and $P=P^{Su}$ on the surface area Su. The strains corresponding to P are

$$\varepsilon=[\varepsilon_{XX}\varepsilon_{YY}\varepsilon_{ZZ}\varepsilon_{XY}\varepsilon_{YZ}\varepsilon_{ZX}], \tag{A2}$$

where $\varepsilon_{XX}=\partial P/\partial X$, $\varepsilon_{YY}=\partial P/\partial Y$, $\varepsilon_{ZZ}=\partial P/\partial Z$, $\varepsilon_{XY}=\partial P/\partial X+\partial P/\partial Y$, $\varepsilon_{YZ}=\partial P/\partial Y+\partial P/\partial Z$, $\varepsilon_{ZX}=\partial P/\partial Z+\partial P/\partial X$. The stresses corresponding to ε are

$$\tau=[\tau_{XX}\tau_{YY}\tau_{ZZ}\tau_{XY}\tau_{YZ}\tau_{ZX}], \tag{A3}$$

where $\tau=C\varepsilon+\tau^I$. C is the stress–strain material matrix and the vector τ^I denotes given initial stresses.

The basis of the displacement-based finite-element solution is the principle of virtual displacements, i.e., the principle of virtual work. This principle states that the equilibrium of the body in Fig. 14 requires that for any compatible small virtual displacements imposed on the body in its state of equilibrium, the total internal virtual work is equal to the total external virtual work

$$\int_V \varepsilon^T\tau\,dV=\int_V P^T f^B\,dV+\int_{s_f} P^{s_f T} f^{s_f} dS+\sum_i P^{iT}R^i, \tag{A4}$$

where P is the virtual displacement and ε is the corresponding virtual strain. When the principle of virtual displacements is satisfied for all virtual displacements with stresses τ, all three fundamental requirements of the mechanics, equilibrium, compatibility, and the stress–strain law are fulfilled. Stresses τ are obtained from a continuous displacement field P that satisfies the displacement boundary conditions on $\mathbf{S_u}$.

In the finite-element analysis, we approximate the body in Fig. 14 as an assemblage of discrete finite elements interconnected at the nodal points on the element boundaries. Displacement p measured in a local coordinate system *x,y,z* within each element is assumed to be a function of the displacement at all nodal points bounded on the element. Therefore, for element m we have

$$p^{(m)}(x,y,z)=H^{(m)}(x,y,z)X^{(m)}, \tag{A5}$$

where $H^{(m)}$ is the displacement interpolation matrix and $X^{(m)}$ is a vector of global displacement components in three dimensions for all nodal points.

Figure 14 shows a typical finite element of the assemblage. This element has eight nodal points, one at each corner, which can be thought of as a ''brick'' element. The complete body is represented as an assemblage of such brick

elements put together so that no gap is left between the element domains. Therefore, for a given brick element, the displacement at the nodes can fully describe the displacement and strain distributions within the element.

With the assumption on the displacements in the above equation, the corresponding element strain is given by

$$\varepsilon^{(m)}(x,y,z) = D^{(m)}(x,y,z)X^{(m)},$$
$$\tau^{(m)}(x,y,z) = C^{(m)}\varepsilon^{(m)}(x,y,z) + \tau^{I(m)}, \tag{A6}$$

where $D^{(m)}$ is the strain-displacement matrix, and the rows of $D^{(m)}$ are obtained by appropriately differentiating and combining the rows of the $H^{(m)}$. $C^{(m)}$ is the stress–strain matrix of the material, which depends only on the Young's modulus and Poisson ratio. An effective assembly process of all matrices into the governing structure matrices can be achieved by using the above equations in the principle of virtual displacement. This process is referred to as *the direct stiffness method*.

Using the direct stiffness method, the stiffness matrix K can be calculated for the element assemblage

$$K = \sum_m \int_{V^{(m)}} D^{(m)^T} C D^{(m)} dV^{(m)}, \tag{A7}$$

where the generalized stress–strain matrix C becomes independent of the elements with the assumption that the material is isotropic. Using d'Alembert's principle, the element inertia forces can be included in the body forces. Thus, the mass matrix M of the structure can be obtained from the integral of (A8). For the velocity-dependent damping forces, the damping matrix B can be obtained in the same way

$$M = \sum_m \int_{V^{(m)}} \rho H^{(m)^T} H^{(m)} dV^{(m)},$$
$$B = \sum_m \int_{V^{(m)}} b H^{(m)^T} H^{(m)} dV^{(m)}, \tag{A8}$$

where ρ is the mass density, and b is the damping property parameter.

Based on the description above, the derived equation of equilibrium governing the linear dynamic response of a finite element system is

$$M\ddot{X} + B\dot{X} + KX = F, \tag{A9}$$

where X, $\dot{X}$, and $\ddot{X}$ are the displacement, velocity, and acceleration vectors of the finite element assemblage, respectively. F is the vector of externally applied loads, described in the following integral:

$$F = \sum_m \left(\int_{V^{(m)}} H^{(m)} f^B \, dV^{(m)} + \int_{S^{(m)}} H^{S(m)} f^S \, dS^{(m)} \right) + f^L, \tag{A10}$$

where f^B, f^S, and f^L represent body, surface, and concentrated forces, respectively.

According to X-FEM, the matrices M, B, and K, and the force F are derived from (A7) and (A8), in which the description is based on the displacements of the nodal points. For a loading force, thus, the displacements X can be obtained from (A9), and then the stress can be evaluated using (A6).

Alfonso, P. J., Watson, B. C., and Baer, T. (**1987**). ''Measuring stutterers' dynamical vocal tract characteristics by x-ray microbeam pellet tracking,'' in *Speech Motor Dynamics in Stuttering*, edited by H. F. M. Peters and W. Hulstijn (Springer, New York).

Baer, T., Alfonso, J., and Honda, K. (**1988**). ''Electromyography of the tongue muscle during vowels in /∂pvp/ environment,'' Ann. Bull. R. I. L. P., Univ. Tokyo **7**, 7–18.

Bathe, K. (**1996**). *Finite Element Procedures* (Prentice-Hall, Englewood Cliffs, NJ).

Belytschko, T., Parimi, C., Moës, N., Sukumar, N., and Usui, S. (**2002**). ''Structured extended finite element methods for solids defined by implicit surfaces,'' Int. J. Numer. Methods Eng. **56**(4), 609–635.

Dang, J., and Honda, K. (**1997**). ''Correspondence between three-dimensional deformation and EMG signals of the tongue,'' Proceedings of the Spring Meeting of Acoust. Soc. Jpn., 241–242 (in Japanese).

Dang, J., and Honda, K. (**1998**). ''Speech production of vowel sequences using a physiological articulatory model,'' Proc. ICSLP98, 5, 1767–1770.

Dang, J., and Honda, K. (**2001**). ''A physiological model of a dynamic vocal tract for speech production,'' Acoust. Sic. Tech **22**, 415–425.

Dang, J., and Honda, K. (**2002**). ''Estimation of vocal tract shape from sounds via a physiological articulatory model,'' J. Phonetics **30**, 511–532.

Dang, J., Honda, M., and Honda, K. (**2002**). ''Investigation of coarticulation based on electromagnetic articulographic data,'' Proc. ICSLP2002, 1521–1524.

Dang, J., Honda, K., and Tohkura, Y. (**1997**). ''3D observation of the tongue articulatory movement for Chinese vowels,'' Technical Report of IEICE, SP97–11, 9–16.

Dang, J., Sun, J., Deng, L., and Honda, K. (**1999**). ''Speech synthesis using a physiological articulatory model with feature-based rules,'' Proc. ICPhS-99, 2267–2270.

Davis, E., Douglas, A., and Stone, M. (**1996**). ''A continuum mechanics representation of tongue motion in speech,'' Proceedings of ICSLP1996, 788–792.

Duck (**1990**). *Physical Property of Tissues: A Comprehensive Reference Book* (Academic, London).

Erickson, D. (**2002**). ''Articulation of extreme formant patterns for emphasized vowels,'' Phonetica **59**, 134–149.

Feldman, A. (**1986**). ''Once more on the equilibrium-point hypothesis (λ model) for motor control,'' J. Motor Behav. **1**, 17–54.

Fung, Y. (**1993**). *Biomechanics—Mechanical Properties of Living Tissue*, 2nd ed. (Spriger, New York).

Hashi, M., Westbury, J., and Honda, K. (**1998**). ''Vowel posture normalization,'' J. Acoust. Soc. Am. **104**, 2426–2437.

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.

Hill, A. V. (**1938**). ''The heat of shortening and the dynamic constants of muscle,'' Proc. R. Soc. London, Ser. B **126**, 136–195.

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

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

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. R. I. L. P. Univ. Tokyo **10**, 243–251.

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

Min, Y., Titze, I., and Alipour, F. (**1994**). ''Stress–strain response of the human vocal ligament,'' NCVS Status Prog. Rep. **7**, 131–137.

Miyawaki, K. (**1974**). ''A study of the muscular of the human tongue,'' Ann. Bull. R. I. L. P. Univ. Tokyo **8**, 23–50.

Morecki, A. (**1987**). ''Modeling, mechanical description, measurements and control of the selected animal and human body manipulation and locomo-

tion movement,'' in *Biomechanics of Engineering—Modeling, Simulation, Control*, edited by A. Morecki (Springer, New York), pp. 1–28.
Napadow, V. J., Chen, Q., Wedeen, V. J., and Gilbert, R. J. (**1999**). ''Intramural mechanics of the human tongue in association with physiological deformations,'' J. Biomech. **32**, 1–12.
Niimi, S., Kumada, M., and Niitsu, M. (**1994**). ''Functions of tongue-related muscles during production of the five Japanese vowels,'' Ann. Bull. R. I. L. P. Univ. Tokyo **28**, 33–40.
Oka, S. (**1974**). *Rheology–Biorheology* (Shokabo, Tokyo, 1974), pp. 445–456.
Okadome, T., and Honda, M. (**2001**). ''Generation of articulatory movements by using a kinematic triphone model,'' J. Acoust. Soc. Am. **109**, 453–463.
Payan, Y., and Perrier, P. (**1997**). ''Synthesis of V–V sequences with a 2D biomechanical tongue shape in vowel production,'' Speech Commun. **22**, 185–206.
Perkell, J. (**1974**). ''A physiological-oriented model of tongue activity in speech production,'' Ph.D. thesis, MIT.
Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). ''Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92**, 3078–3096.
Sanguineti, V., Laboissière, J., and Payan, Y. (**1997**). ''A control model of human tongue movements in speech,'' Biol. Cybern. **77**, 11–22.
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.
Stevens, K. (**2000**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
Stone, M., Davis, E., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W., and Lundberg, A. (**2001**). ''Modeling motion of the internal tongue from tagged cine-MRI images,'' J. Acoust. Soc. Am. **109**, 2974–2982.
Takemoto, H. (**2001**). ''Morphological analyses of the human tongue musculature for three-dimensional modeling,'' J. Speech Lang. Hear. Res. **44**, 95–107.
Warfel, J. (**1993**). *The Head, Neck, and Trunk* (Led & Febiger, Philadelphia and London).
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3805–3898.
Yamazaki, K. (**1933**). ''The weight of the cranium and mandible with comparison of the dental and bony regions of the mandible,'' Jpn. J. Dentistry **26**, 769–796 (in Japanese).
Zajac, F. (**1989**). ''Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control,'' Crit. Rev. Biomech. Eng. **17**, 359–411.
Zienkiewicz, O., and Taylor, R. (**1989**). *The Finite Element Method* (McGraw-Hill, New York).