Acoust. Sci. & Tech. **35**, 4 (2014)

PAPER

Iterative method to estimate muscle activation with a physiological articulatory model

Xiyu Wu[1,*], Jianwu Dang[1,2,†] and Ian Stavness[3,‡]

[1]*Japan Advanced Institute of Science and Technology, 1–1 Asahidai, Nomi, 923–1292 Japan*
[2]*Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, 92 Weijin Road, Nankai District, Tianjin, 300072, P. R. China*
[3]*Department of Computer Science, University of Saskatchewan, 176 Thorvaldson Building, 110 Science Place, Saskatoon SK S7N 5C9, Canada*

(*Received 20 May 2013, Accepted for publication 21 January 2014*)

Abstract: Computational modeling of the speech organs is able to improve our understanding of human speech motor control. In order to investigate muscle activation in speech motor control, we have developed an automatic estimation method based on a 3D physiological articulatory model. In this method, the articulatory target was defined by the entire posture of the tongue and jaw in the midsagittal plane, which was reduced to a six-dimensional space by principal component analysis (PCA). In the PCA space, the distance between an articulatory target and the model was gradually minimized by automatically adjusting muscle activations. The adjustment of muscle activations was guided by a dynamic PCA workspace that was used to predict individual muscle functions in a given position. This dynamic PCA workspace was estimated on the basis of an interpolation of eight reference PCA workspaces. The proposed method was assessed by estimating muscle activations for five Japanese vowel postures that were extracted from magnetic resonance images. The results showed that the proposed method can generate muscle activation patterns that can control the model to realize given articulatory targets. In addition, the estimated muscle activation patterns were consistent with anatomical knowledge and previously reported measurement data.

Keywords: Physiological articulatory model, Muscle activation, Speech production, Motor control

PACS number: 43.70.Bk, 43.70.Jt [doi:10.1250/ast.35.201]

1. INTRODUCTION

Speech organs are driven by coordinated muscle activations to produce vocal tract shapes during speech production. In order to improve the understanding of speech motor control, many physiological articulatory models have been constructed since the 1970s [1–5]. To investigate muscle activations with a physiological articulatory model, Fang *et al.* [4] and Buchaillard *et al.* [5] used the "trial-and-error" method, where muscle activation patterns were obtained by manually adjusting muscle activations to minimize the distance between the model simulation and target posture. Because the "trial-and-error" method depends on the experiential knowledge of the researcher, it is difficult to estimate muscle activation patterns for specific postures. To avoid the disadvantages of the "trial-and-error" method, some automatic strategies have been proposed to estimate muscle activation patterns. Dang and Honda constructed a set of maps from the equilibrium position of three control points (tongue tip, tongue dorsum, and jaw) to muscle activations and used these maps to estimate muscle activations according to the target of the control points [3]. Stavness *et al.* estimated muscle activations from movements of the tongue tip [6]. However, in speech production, the phonetic qualities of speech sounds depend on the whole vocal tract shape rather than only the size and location of the vocal tract constriction at the tongue tip or tongue dorsum. Therefore, the automatic estimation of muscle activations for a given articulatory posture is necessary for exploring speech motor control.

Articulatory organs that combine musculoskeletal (jaw) and muscular-hydrostat (tongue) structures are complex

*e-mail: xiyuwu@jaist.ac.jp
†e-mail: jdang@jaist.ac.jp
‡e-mail: stavness@gmail.com

biomechanical systems. For musculoskeletal systems, inverse estimation techniques have been applied to automatically predict muscle activations for prescribed kinematics, including lower limb [7,8] and hand [9] movements. For muscular-hydrostat systems, Stavness *et al.* predicted coordination of muscle activations using forward-dynamics tracking simulation [6]. Estimating muscle activations for articulatory postures that are shaped by muscular-hydrostat and musculoskeletal systems is difficult because 1) coupling effects cause the activation of one muscle to affect not only a specific component but also the whole system, and 2) different muscle activation patterns may generate the same articulatory posture (the "one-to-many" problem).

The purpose of this study is to develop a method to automatically estimate muscle activations according to given articulatory postures. To do so, we first describe our physiological articulatory model, which has been upgraded from discrete FEM to continuum FEM. We then describe the proposed method to estimate muscle activations using the model. Finally, we report an evaluation of muscle activation patterns estimated for five Japanese vowel postures extracted from magnetic resonance images.

2. PHYSIOLOGICAL ARTICULATORY MODEL

The original version of the physiological articulatory model was a partial 3D model constructed by Dang and Honda and was based on discrete FEM [3]. This model was developed to a full 3D model by Fujita *et al.* [10]. In this study, the physiological articulatory model was extended to a continuum FE model using the ArtiSynth 3D Biomechanical Modeling Toolkit (www.artisynth.org, University of British Columbia, Vancouver, Canada). ArtiSynth improved a number of aspects of the physiological articulatory model, including volume constraint and computational efficiency. The profile of the constructed model is shown in Fig. 1, where the appearance of the model is shown in the left panel and a sagittal cut-away view is shown in the right panel.

2.1. Dynamic Simulation

The ArtiSynth toolkit was used to generate dynamic simulations with the physiological articulatory model. In this section, we describe the pertinent equations and parameters used in this physiological articulatory model. A complete description of the dynamic simulation formulation in ArtiSynth has been published elsewhere (see Section 4 in [11]).

According to Newton's second law, the equations of motion that govern the dynamic response of the finite element system are given by

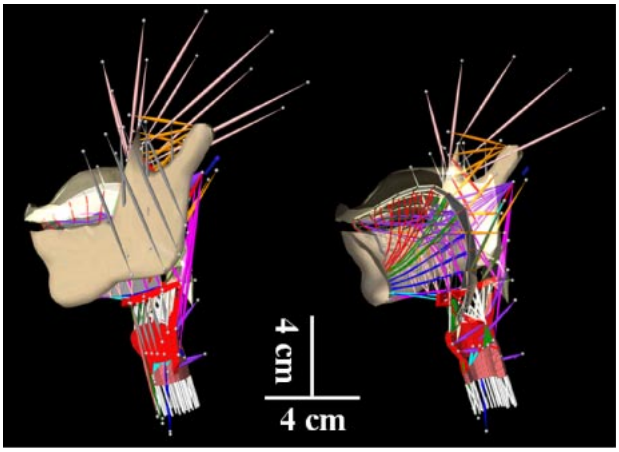

Fig. 1 Lateral (left) and mid-sagittal cutaway (right) views of the physiological articulatory model.

$$\boldsymbol{M}\dot{\boldsymbol{u}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{u}, t), \tag{1}$$

where t is time, $\boldsymbol{q}$ and $\boldsymbol{u}$ are the generalized position and velocity of all dynamical components in the mechanical system, $\boldsymbol{f}(\boldsymbol{q}, \boldsymbol{u}, t)$ is the total forces acting on the dynamic components, and $\boldsymbol{M}$ is the block-diagonal mass matrix.

The system dynamics are also constrained by bilateral and unilateral constraints. Bilateral constraints are used to attach the tongue to the jaw and hyoid bone, as well as to enforce FEM incompressibility in the FE models (through a mixed u-P formulation [12]). Bilateral constraints form an equality condition on the system velocity $\boldsymbol{u}$:

$$\boldsymbol{G}(\boldsymbol{q})\boldsymbol{u} = 0. \tag{2}$$

Unilateral constraints are used to handle contact between the tongue tip, the jaw, and palate. Unilateral constraints form an inequality condition on the system velocity $\boldsymbol{u}$:

$$\boldsymbol{N}(\boldsymbol{q})\boldsymbol{u} \geq 0. \tag{3}$$

Bilateral and unilateral constraints generate reaction forces $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\lambda}$ and $\boldsymbol{N}^{\mathrm{T}}\boldsymbol{z}$, respectively, where $\boldsymbol{\lambda}$ and $\boldsymbol{z}$ are the Lagrange multipliers. These reaction forces add to the system forces in Eq. (1).

The dynamic equations are solved numerically using a semi-implicit second-order Newmark integrator [13]. The update rules for this integration scheme are

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^{k} + \frac{h}{2}(\dot{\boldsymbol{u}}^{k} + \dot{\boldsymbol{u}}^{k+1}), \tag{4}$$

and

$$\boldsymbol{q}^{k+1} = \boldsymbol{q}^{k} + \frac{h}{2}(\boldsymbol{u}^{k} + \boldsymbol{u}^{k+1}), \tag{5}$$

where h is the time step.

Solving the equations of motion requires integrating Eq. (1) with the update steps given in Eqs. (4) and (5)

subject to the conditions given in Eqs. (2) and (3). This requires solving the following mixed linear complementarity problem:

$$\begin{pmatrix} \hat{\boldsymbol{M}}^k & -\boldsymbol{G}^{k\mathrm{T}} & -\boldsymbol{N}^{k\mathrm{T}} \\ \boldsymbol{G}^k & 0 & 0 \\ \boldsymbol{N}^k & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^{k+1} \\ \boldsymbol{\lambda} \\ \boldsymbol{z} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{b} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \boldsymbol{w} \end{pmatrix},$$
$$0 \leq \boldsymbol{z} \perp \boldsymbol{w} \geq 0, \tag{6}$$

where $\boldsymbol{b} \equiv \boldsymbol{M}\boldsymbol{u}^k + h\hat{\boldsymbol{f}}^k$, $\boldsymbol{w}$ is the slack variable under the complementarity condition, and $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{f}}$ are the mass matrix and force vector augmented with Jacobian terms due to the implicit integration scheme (see [11] for a full derivation). The complementarity condition, $0 \leq \boldsymbol{z} \perp \boldsymbol{w} \geq 0$, ensures that the unilateral constraint forces are nonzero only when those constraints are active, i.e., $\boldsymbol{z}$ is positive if and only if $\boldsymbol{w}$ is zero and *vice versa*.

2.2. Model Structure

The morphological structures of the tongue, the jaw, and the vocal tract wall were extracted from magnetic resonance (MR) images. The jaw and the vocal tract wall were superimposed with the images of the lower and upper teeth at intervals of 0.4 cm in the transverse dimension. The initial shape of the tongue was obtained from the volumetric MR images taken while producing the Japanese vowel /e/, which is close to the neural position in vowel space. The mesh structure of the tongue in the lateral view consists of eleven layers with nearly equal intervals fanning out to the tongue surface from the attachment on the mandible, and seven layers in the perpendicular direction. In the front view, the tongue was divided into 5 layers at equal intervals. Totally, the tongue tissue consists of 240 hexahedrons. For more details of morphological data and mesh segmentation of the tongue tissue, refer to the previous studies [4,10].

2.3. Mechanical Properties

In the present physiological articulatory model, Rayleigh damping was implemented in the form

$$\boldsymbol{D}_{\mathrm{F}} = \alpha \boldsymbol{M}_{\mathrm{F}} + \beta \boldsymbol{K}_{\mathrm{F}}, \tag{7}$$

where $\boldsymbol{M}_{\mathrm{F}}$ is the portion of the mass matrix associated with the FEM nodes and $\boldsymbol{K}_{\mathrm{F}}$ is the FEM stiffness matrix. $\boldsymbol{D}_{\mathrm{F}}$ is embedded into the overall system Eq. (6) by $\boldsymbol{D}_{\mathrm{F}} = \partial \boldsymbol{f} / \partial \boldsymbol{u}$. In the present model, α and β were set to $40\,\mathrm{s}^{-1}$ and 0.03 s, respectively, in order for a damping to be close to the critical one in the range of modal frequency from 3 to 10 Hz [5]. For details on how to integrate Rayleigh damping into Eq. (6), refer to paper the paper by Stavness *et al.* [14].

A Poisson coefficient of tongue tissue was set to 0.49 since it was considered to be quasi-incompressible. The density of tongue tissue was set to be $1{,}040\,\mathrm{kg\,m^{-3}}$, and the density of the mandible and hyoid bone was set at $2{,}000\,\mathrm{kg\,m^{-3}}$. Young's modulus of the tongue tissue was set at 20 kPa and the bone structures (mandible and hyoid) were approximated as rigid bodies. These parameters were consistent to those of the previous model [3].

2.4. Muscle Structures

Since the model is driven by muscle activation, the accuracy of muscle implemention is very important. Three extrinsic muscles, the genioglossus, styloglossus, and hyoglossus were arranged mainly on the basis of the results of high-resolution MRI analysis [15]. The intrinsic muscles (superior longitudinal, inferior longitudinal, transverse, and vertical) were defined according to the anatomical data [16]. The tongue floor muscles, mylohyoid and geniohyoid, were arranged referring to the anatomical literature [17]. The muscles that control jaw movements were defined in the same way as that in our previous partial 3D model [3].

The muscle model implemented in this study was proposed by Morecki [18] and is an extended model of Hill's model [19]. Forces generated by muscles include two components: active muscle force that depends on muscle activation and passive muscle force that is independent of muscle activation.

In model computations, the active stress of the muscle sarcomere was generated using the force-length function, which was derived by matching the simulation and empirical data by the least-squares method [18]. In this function, shown in Eq. (8), the active stress (σ_{act}) was calculated using a fourth-order polynomial of the stretch ratio of the muscles, which had a similar shape to that used by Wilhelms-Tricarico [20]. In Eq. (8), the muscle length change rate $\varepsilon = (l - l_0)/l_0$ was valid for the range of $-0.185 < \varepsilon < 0.49$, where l and l_0 are the present muscle length and original muscle length, respectively. Therefore, the active force was set to 0 when ε was out of the given range.

$$\sigma_{\mathrm{act}} = 1.161\varepsilon^4 + 0.243\varepsilon^3 - 1.376\varepsilon^2 + 0.235\varepsilon + 0.164 \tag{8}$$

The ability to generate muscle force varies from muscle fiber to muscle fiber depending on their thickness. Therefore, the parameter "thickness" of the muscle fiber was introduced as a coefficient for all the muscles, and the thickness governs the capacity of force generation. The thickness of individual muscles was determined by making the maximum force ($F_{\max}$) of the muscles consistent with empirical data [21,22]. The control variable of individual muscle activation, a, was normalized within the interval $[0, 1]$, where 0 means no muscle activation and 1 means that the muscle is fully activated and generates maximum force. The activated muscle force was calculated as

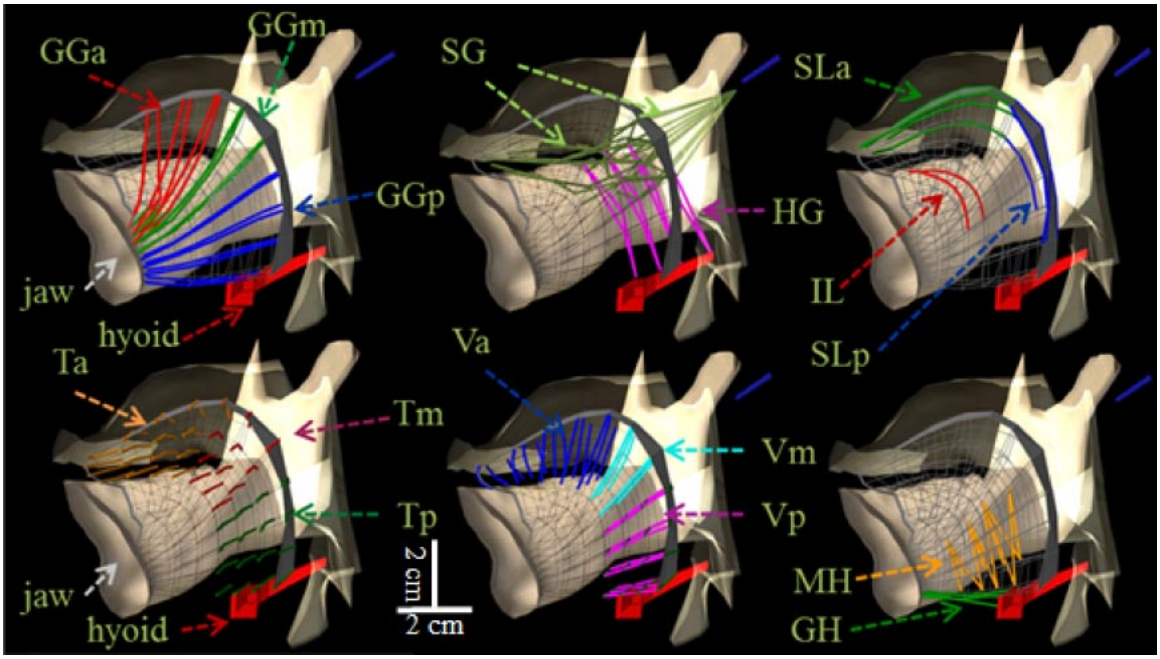

Fig. 2 Arrangement of muscles in the tongue model. GGa, GGm and GGp: anterior, middle, and posterior portions of genioglossus muscle, respectively; HG: hyoglossus muscle; SG: styloglossus muscle; SLa and SLp: the anterior and posterior portions of the superior longitudinal muscle; IL: inferior longitudinal muscle; Va, Vm and Vp: anterior, middle, and posterior portions of vertical muscle, respectively. Ta, Tm and Tp: anterior, middle, and posterior portions of transverse muscle, respectively; MH: mylohyoid muscle; GH: geniohyoid muscle.

$$F_{\text{act}} = F_{\max}\sigma_{\text{act}}a, \tag{9}$$

where $F_{\max}$ is the maximum isometric force capacity of the muscle, σ_{act} is the active muscle stress (see Eq. (8)), and a is muscle activation.

Passive muscle force is generated by passive lengthening of muscle. According to common sense, if a muscle is lengthened to become equal to or longer than a threshold $l_{\text{p\_max}}$, the passive muscle force will no longer continue increasing with lengthening and reach the maximum passive force ($F_{\text{p\_max}}$) that the muscle can generate. In the present physiological articulatory model, $l_{\text{p\_max}}$ was set to 1.25 times the original muscle length and $F_{\text{p\_max}}$ was related to $F_{\max}$ by $F_{\text{p\_max}} = 0.015F_{\max}$ according to Morecki's muscle model [18]. The passive muscle force is described by

$$F_{\text{pas}} = \begin{cases} 0 & \text{if } l < l_0 \\ F_{\text{p\_max}}[(l - l_0)/(l_{\text{p\_max}} - l_0)] & \text{if } l_0 < l < l_{\text{p\_max}} \\ F_{\text{p\_max}} & \text{if } l \geq l_{\text{p\_max}}. \end{cases} \tag{10}$$

The final muscle force was the sum of active muscle force F_{act} and passive muscle force F_{pas}.

2.5. Muscle Units for Model Control

Muscles in the model were arranged on the basis of their anatomical partitions where different parts of the same muscle may have different functions. In order to simulate fine-grained tongue movements with the model, the muscles were divided into a number of smaller control units according to articulation purposes. Figure 2 illustrates the layout of the extrinsic and intrinsic muscles (original or divided) in the 3D physiological articulatory model in a sagittal cut-away view. The genioglossus muscle was divided into three units: anterior (GGa), middle (GGm), and posterior (GGp). This division conforms to those of previous physiological articulatory models [3–5]. Different from the previous studies [3–5], the intrinsic muscles were also divided into several control units according to their functions. The vertical and transverse muscles were functionally divided into three units: anterior (Va, Ta), middle (Vm, Tm), and posterior (Vp, Tp). The superior longitudinal was divided into two units: anterior (SLa) and posterior (SLp). The styloglossus (SG), mylohyoid (MH), geniohyoid (GH), and inferior longitudinal (IL) were controlled as independent units.

The muscles used to control the translation and rotation of the jaw were classified into two muscle groups: the jaw opener (JO) and jaw closer (JC). In Fig. 3, the arrangements of muscles used to control the jaw are described. According to the description of the muscles used to control the jaw [23], the jaw opening muscles include the anterior digastrics, posterior digastrics, and lateral pterygoid muscles. The strap muscle sternohyoid also assists jaw opening. The main function of the lateral pterygoid is to move the jaw forward, but the current version of the jaw model only

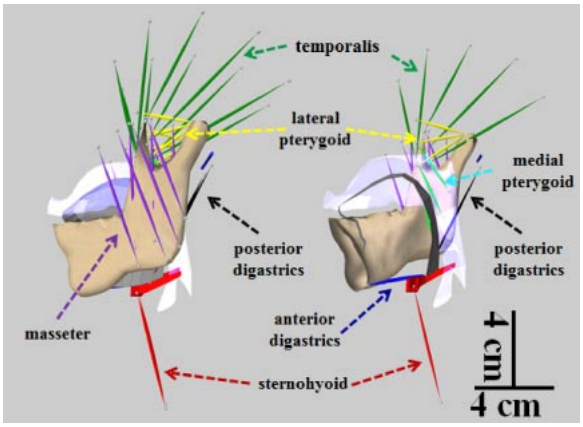

Fig. 3 Arrangement of muscles in the jaw model.

permits hinge-like jaw opening. Therefore, in this study, the JO group consists of the anterior digastrics, posterior digastrics, and sternohyoid. When JO was activated, the muscles in the group were active with the same activation level. The jaw closing muscles include the temporalis, masseter, and median pterygoid muscles. Among those muscles, comparatively small muscles are used for speech articulation, while larger muscles play major roles in biting and chewing [23]. The medial pterygoid plays the main role in speech production, while the temporalis and masseter contribute less. According to our simulation, the activation level for the temporalis and masseter were the fourth and fifth of that for the medial pterygoid, respectively. To control the physiological articulatory model at a higher degree of freedom, some muscles were divided into smaller units, while some muscles were combined into groups. Altogether, 18 muscle control units were used in this study.

2.6. Muscle Functions

The activity of each muscle contributes to local deformation or displacement of the articulatory organs. The estimation of muscle activation patterns relies on the function of individual muscle units. For this reason, we investigated the functions of the muscle units individually. In the simulations, each muscle was activated individually for the duration of 200 ms, which was sufficient for the model to reach its equilibrium position. The functions of the extrinsic and intrinsic muscles were qualitatively assessed on the basis of anatomical description. These assessments show that the role of individual muscles in our model was consistent with anatomical knowledge [24–26]. The difference in muscle control units between the present and previous models [4,5] was that, in the present model some intrinsic individual muscles were divided into smaller control units according to their functions. Figure 4 shows the functions of some intrinsic individual muscle units on the midsagittal plane. From Fig. 4, one can see that different portions of the vertical muscles have different functions (refer to the functions of Va, Vm and Vp). Similarly, the control units Ta, Tm and Tp have different functions although they belong to the same muscle (transverse muscle). These imply that the divisions of the muscle units were effective.

There is a common feature of FEM-based physiological articulatory models: when a muscle activation pattern is maintained, the model reaches a certain equilibrium position. This equilibrium position is determined only by

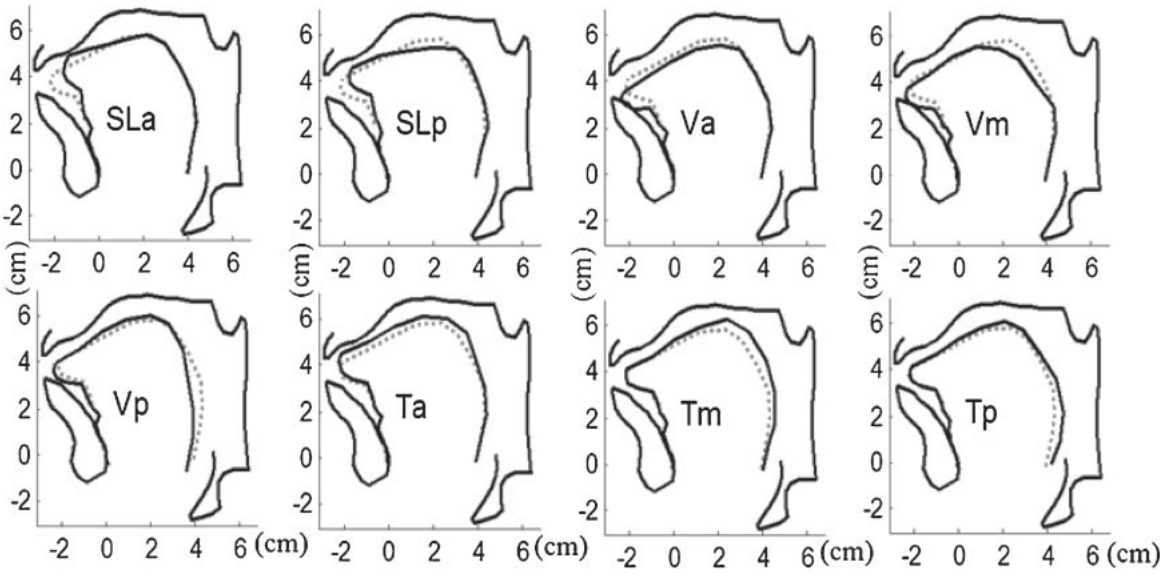

Fig. 4 Function of part of individual muscles in the physiological articulatory model. Black solid lines show the equilibrium position after the muscle is activated for a 200 ms duration, dotted gray lines correspond to the shape in its rest position. Horizontal and vertical axes are 'anteroposterior' and 'vertical,' respectively.

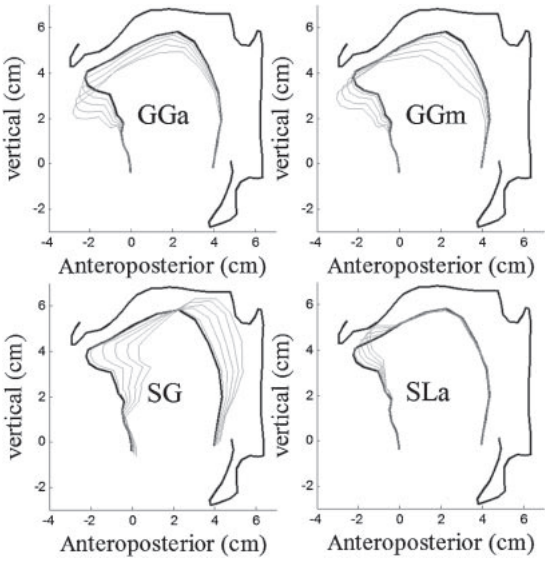

Fig. 5 Muscle activation and equilibrium position. Black thick lines are the rest position of the tongue; gray lines are the equilibrium position driven at different activation levels [0.002, 0.01, 0.03, 0.1, 0.4].

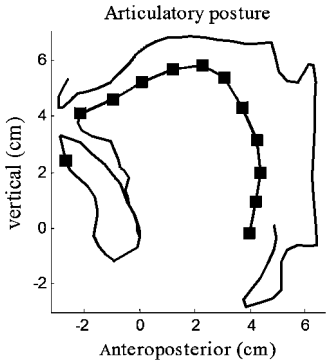

Fig. 6 Representation of articulatory posture in the midsagittal plane.

Table 1 Variance of PCA components (%). C1 to C6 are the first six components, VC is the variance of the components, and AVC is the accumulative variance from the first component to the current component.

Component	C1	C2	C3	C4	C5	C6
VC	79.58	12.54	3.01	2.32	1.42	0.46
AVC	79.58	92.12	95.13	97.45	98.87	99.33

muscle activation itself, no matter where its initial position is. Figure 5 shows the changes in the equilibrium position when the activation level is changed. In this figure, GGa, GGm, SG, and SLa are active with the activation levels of 0.002, 0.01, 0.03, 0.1 and 0.4. After 200 ms, the model reaches its equilibrium position. From this figure, one can see that the muscle activation level determines a unique equilibrium position.

Previous model simulations have shown that the relationship between muscle activation level and displacement of the tongue is quasi-logarithmic [3]. To generate displacements with approximately the same increment, muscle activation was discretized into 11 levels of 0, 0.002, 0.005, 0.01, 0.02, 0.03, 0.05, 0.1, 0.2, 0.4 and 0.8. Section 3.1 provides details on the chosen activation levels.

3. MUSCLE ACTIVATION ESTIMATION METHOD

In the previous studies, articulatory targets were defined by isolated control points (tongue tip, tongue dorsum, and jaw), and these points were used to control the constriction position of the vocal tract [3,27]. Since the acoustic characteristics of speech sounds depend on the whole vocal tract configuration, the contour of the tongue and jaw is a proper target for model control. In this study, we use the midsagittal contour to describe the articulatory posture, which can represent most the phonemes, except some lateral ones. In addition, it is convenient to measure articulatory movement on the midsagittal plane since this is the output of many common observation techniques, such as electromagnetic articulography (EMA), X-ray microbeam, and MR imaging. Therefore, the articulatory posture defined in the midsagittal plane will facilitate the comparison of model shapes with the measurement data. Consequently, eleven points on the tongue surface and one point on the lower incisor are used to represent the articulatory posture of our model, as seen in Fig. 6.

3.1. Principal Component Analysis of Articulatory Posture

As described previously, each articulatory posture is depicted by 12 points with the horizontal and vertical coordinates. However, these points have significant redundancy and correlativity. To reduce redundancy, principal component analysis (PCA) was adopted to analyze the posture patterns of the model in the midsagittal plane.

To generate a data set for PCA, our objective was to create simulations that cover most of the possible postures by using reasonable muscle combinations considering the agonist-antagonist properties of muscles. With reference to the previous study [28] concerning the agonist-antagonist muscles and muscle combinations, 9,703 articulatory postures that cover most of the possible postures were obtained for PCA.

The variance of each component and the accumulated explanations of variance are shown in Table 1. From this table, one can see that the first six components can explain

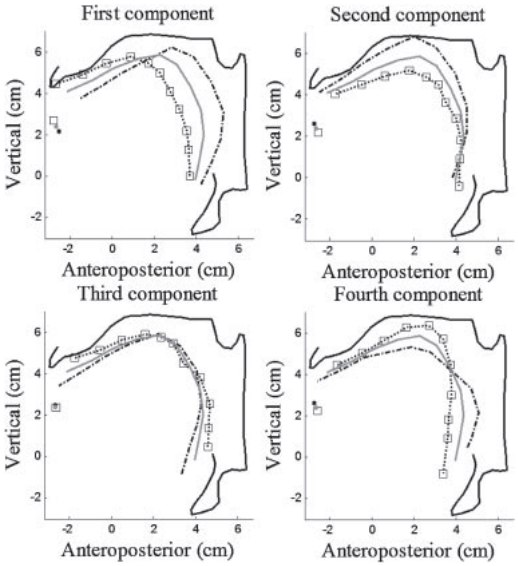

Fig. 7 Effect of the PCA components. Gray lines show the rest position, dashed lines and dashed lines with squares show the directions of each component with positive and negative coefficients, respectively.

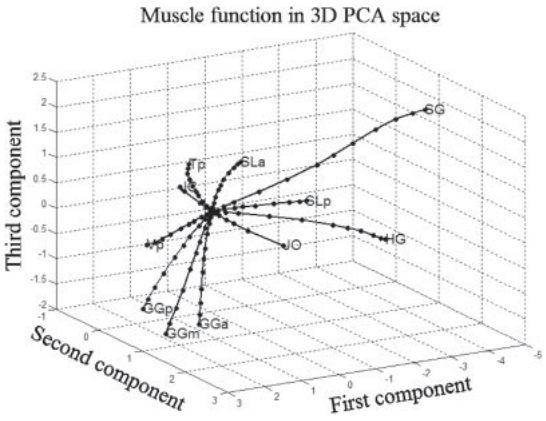

Fig. 8 Function of individual muscle units in 3D PCA space.

99.33% of the variance, which indicates that the articulatory posture can be determined by the first six components within 0.7% error. The contribution of the first four components is shown in Fig. 7. Figure 8 shows the functions of individual muscles by transforming the articulatory postures, which resulted from the individual muscle activation, into the PCA space consisting of the first three components, where some muscles with small impact are not shown. In this figure, each curve shows the function of a single muscle unit in PCA space, where the dots on the curves indicate the results using different muscle activation levels. From this figure, one can see that the equilibrium positions shift from the rest position in PCA space as the activation level increases. There is a monotonic relationship between the muscle activation level and displacement increment: increasing the activation level will drive the articulator to move away from the rest position, whereas decreasing the activation level will make the model return towards the rest position. This monotonicity is crucial for the estimation of muscle activation because we can increase activation if the realized position does not reach the target and decrease the activation if the realized position exceeds the target.

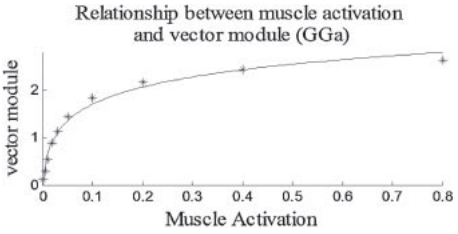

Fig. 9 Relationship between muscle activation and PCA vector module (GGa).

In the PCA space, the relationship between the module of the six-dimensional PCA vector and the activation of individual muscle is quasi-logarithmic according to the results of simulations. This relationship can be represented by the fitted curve described by Eq. (11), where x and y are two undetermined coefficients, a is the activation, and ML is the module of the PCA vector. In order to calculate the undetermined coefficients of Eq. (11) for each muscle, the individual muscles were activated at 20 equal intervals between 0 and 1 to obtain the corresponding data pair of muscle activation and PCA vector module. According to these results, the undetermined coefficients were obtained for individual muscles. Figure 9 shows a fitted curve for GGa muscle, where $x = 0.53501$ and $y = 230.8211$. For each muscle, 1/10 of the PCA vector module generated by 0.8 activation was defined as a scale unit. The increase in muscle activation that causes the vector module to increase by one scale unit is defined as a *unit increment* of muscle activation. The *unit increment* of muscle activation for each muscle can be calculated using Eq. (11). In Fig. 9, stars show the ten *unit increments* of muscle activation and their PCA modules of GGa muscle.

$$ML = x \log_e(ya + 1) \tag{11}$$

3.2. Procedure of Muscle Estimation

An iteration method was used to find muscle activation patterns by gradually minimizing the difference between the target posture and realized position. The distance (D) between the target posture and realized posture was calculated using Eq. (12), where Rx_p and Ry_p were the horizontal and vertical coordinate values of the pth point used to represent the realized posture, and Tx_p and Ty_p were the coordinate values of the corresponding target points.

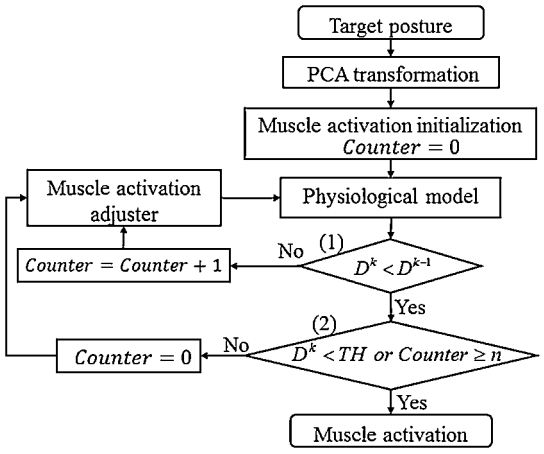

Fig. 10 Flowchart of muscle activation estimation.

$$D = \frac{1}{12}\sum_{p=1}^{12}\sqrt{(Rx_p - Tx_p)^2 + (Ry_p - Ty_p)^2} \qquad (12)$$

The flowchart of the estimation procedure is shown in Fig. 10. A target posture is projected into the 6-dimensional PCA space described above in order to obtain a PCA vector. The muscle activation pattern is initialized using the difference between the rest posture and target posture, and the *Counter* used to count the failed iteration times is initialized to 0. The muscle activation initiation method will be introduced in Sect. 3.3. The muscle activation is input to the physiological articulatory model, and then the model moves to a certain position and reaches equilibrium. If the distance in the kth iteration D^k is smaller than the distance generated from the previous iteration D^{k-1}, the muscle activation is accepted and we move on to decision (2), otherwise we increase the failed counter and go to the *Muscle Activation Adjuster* module. In decision (2), we output the muscle activation if the distance is smaller than the threshold ($D^k < TH$) or the *Counter* is greater than the number of muscle units $Counter \geq n = 18$; if the output conditions are not satisfied, set the *Counter* to 0 and go to the *Muscle Activation Adjuster* module. The most important module in this procedure is the *Muscle Activation Adjuster*, which will be introduced in the next section.

3.3. Dynamic PCA Workspace

In each iteration step, the muscle activation is adjusted by

$$\boldsymbol{a}^k = \boldsymbol{a}^{k-1} + \boldsymbol{\Delta a}^k, \qquad (13)$$

where $\boldsymbol{a}^k$ and $\boldsymbol{a}^{k-1}$ are the muscle activations used in current and previous iteration steps, respectively. The muscle activation vector $\boldsymbol{a} \equiv [a_1 a_2 \ldots a_n]^{\mathrm{T}}$ is constituted by the activation of individual muscle a_i and the number of muscle units $n = 18$.

The main work in each iteration step is to find the adjustive muscle vector $\boldsymbol{\Delta a}^k \equiv [\Delta a_1^k\ \Delta a_2^k \ldots \Delta a_n^k]^{\mathrm{T}}$. Δa_i^k is related to the contribution of individual muscle function vectors to the target vector C_i^k as follows:

$$C_i^k = |\boldsymbol{V}_{mi}^k| \cos\theta = |\boldsymbol{V}_{mi}^k| \frac{\boldsymbol{V}_{mi}^k \cdot \boldsymbol{V}_t^k}{|\boldsymbol{V}_{mi}^k||\boldsymbol{V}_t^k|} = \frac{\boldsymbol{V}_{mi}^k \cdot \boldsymbol{V}_t^k}{|\boldsymbol{V}_t^k|}. \qquad (14)$$

C_i^k is calculated by projecting individual muscle function vectors to the target vector, where $\boldsymbol{V}_{mi}^k$ is the individual muscle function vector of the ith muscle, $\boldsymbol{V}_t^k$ is the target vector and θ is the angle between $\boldsymbol{V}_{mi}^k$ and $\boldsymbol{V}_t^k$. Target vector $\boldsymbol{V}_t^k$ is defined as

$$\boldsymbol{V}_t^k = \boldsymbol{P}_t - \boldsymbol{P}_r^{k-1}, \qquad (15)$$

where $\boldsymbol{P}_t$ is the target posture and $\boldsymbol{P}_r^{k-1}$ is the realized posture after the previous iteration.

The muscle function vector $\boldsymbol{V}_{mi}^k$ is defined by the effect of the activation of the ith muscle with a *unit increment*:

$$\boldsymbol{V}_{mi}^k = \boldsymbol{P}_{i+1} - \boldsymbol{P}_r^{k-1}, \qquad (16)$$

where $\boldsymbol{P}_r^{k-1}$ is the posture realized by activation $\boldsymbol{a}^{k-1}$, and $\boldsymbol{P}_{i+1}$ is the posture realized by a *unit increment* of muscle activation for the ith muscle in $\boldsymbol{a}^{k-1}$. *Unit increment* was explained in Sect. 3.1.

It should be noted that the posture used here is defined by a six-dimensional PCA vector. The ith muscle ($i = \arg\max_i(C_i^k)$) with the maximum contribution $C_{\max}^k = \max(C_i^k)$, will have a *unit increment*. For the other muscles, the increased activations are less than a *unit increment* and their proportion to the *unit increment* is calculated by $(C_i^k / C_{\max}^k)$. The increment of muscle activation of a *unit increment* is calculated by the constructed fitting curve for individual muscle using Eq. (11). Note that, after adding $\boldsymbol{\Delta a}^k$ to $\boldsymbol{a}^{k-1}$ in Eq. (13), if the activation of muscle a_i^k is smaller than 0, it will be set to 0, because there is no negative muscle activation.

In the rest position, the individual muscle function vector $\boldsymbol{V}_{mi}^k$ was built by activating individual muscles, as shown in Fig. 8. However, during articulation, muscle orientations vary along with the movement of the jaw and tongue, which will result in the variation of the muscle function vector. To solve this problem, Dang and Honda [27] proposed a method for estimating the muscle function orientation dynamically. Following their idea, we constructed a set of reference PCA workspaces in some extreme locations by moving the origin to given extreme locations. When speech organs move to an arbitrary position, a dynamic PCA workspace can be interpolated on the basis of the reference PCA workspace.

We first construct seven reference PCA (r-PCA) workspaces by the following procedures: 1) move the PCA center to seven extreme locations in PCA space by a set of selected muscle activation patterns; 2) in a given

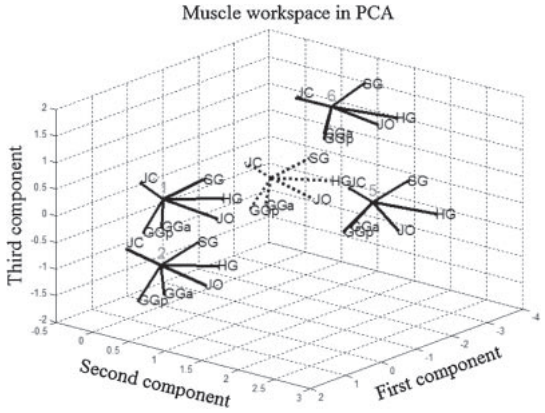

Fig. 11 Reference PCA workspaces (solid lines) and dynamic PCA workspace (dash lines) in 3D PCA.

PCA center, construct a r-PCA workspace by a *unit increment* of individual muscle (refer to Sect. 3.1). Together with the r-PCA workspace in the rest position, we have eight r-PCA workspaces. The r-PCA workspaces in 3D PCA are shown in Fig. 11, where No. 1 is the original PCA workspace in the rest position, and Nos. 2–8 show the other r-PCA workspaces in different reference positions. In order to show the r-PCA workspace clearly, only four r-PCA workspaces with extrinsic muscles are shown in this figure.

The dynamic PCA workspace (d-PCA) for a given position is interpolated from their distance to the eight reference PCA workspace using the following equation:

$$\boldsymbol{V}_{mi}^{k} = \frac{\sum_{s=1}^{w} L_s \boldsymbol{V}_{si}}{\sum_{s=1}^{w} L_s}; \quad L_s = \prod_{\substack{j=1 \\ j \neq s}}^{w} l_j^2, \tag{17}$$

where $\boldsymbol{V}_{mi}^{k}$ denotes a muscle function vector in d-PCA, $\boldsymbol{V}_{si}$ is the muscle function vector in the sth r-PCA workspace, l_j is the Euclidean distance from the current position to the origin of the jth r-PCA workspace, and $w = 8$ is the number of reference PCA workspaces. The coefficient L_s of the sth r-PCA is the product of the distance from the current position to the origin of the other $(w-1)$ r-PCA workspace. The characteristic of the interpolation method is shown in Fig. 12, and demonstrates that the interpolation has a quadratic surface with a relatively flat characteristic surrounding the reference points. Figure 11 shows an example of the d-PCA workspace in the dash lines, which was generated by the interpolation method. The d-PCA workspace reflects the individual muscle function vector in the current position. Occasionally, when the addition of the adjustive muscle vector $\boldsymbol{\Delta a}^k$ cannot control the model to be closer to the target, the adjusted vector $\boldsymbol{\Delta a}^k = [\Delta a_1^k \ \Delta a_2^k \ \ldots \ \Delta a_n^k]$ will be adjusted by setting Δa_i^k to 0, where $|\Delta a_i^k|$ is the smallest nonzero value in the vector, and the *Counter* in Fig. 10 will be increased by 1.

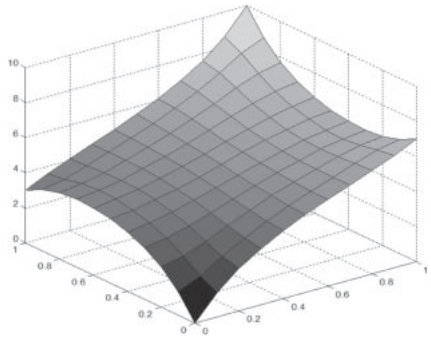

Fig. 12 Example of the interpolation surface using four reference points with coordinates $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ and their values 0, 2.5, 7.5, 10.

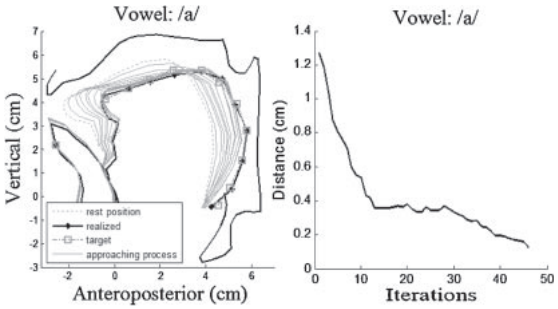

Fig. 13 Processes of muscle activation estimation of vowel /a/.

4. EVALUATION

The proposed method was evaluated using the five Japanese vowel postures obtained from magnetic resonance images as the targets to estimate muscle activation patterns. Since the prototype subject of the physiological model is the same as that for obtaining the MRI data, we can compare them directly without any registration or normalization procedure. In order to control the model to achieve the best target possible, the threshold *TH* shown in Fig. 10 was set to 0 in this experiment. In the iteration process the muscle activations were obtained using the dynamic PCA workspace introduced previously. An example of the iteration processes approaching the target is shown in Fig. 13, where the left panel shows the process of approaching the target, and the right panel is the distance (defined in Eq. (12)) between the target and realized posture during optimization of the muscle activation pattern. In the figure, one can see that the general tendency of the error curve decreases as the iteration increases, and the average distance decreased to 0.123 cm for vowel /a/ after 46 iterations. Although, it is seen that some ripples appeared along with the distance curve, the muscle-adjusting method can modify the muscle activation patterns automatically, and eventually control the model to achieve the target.

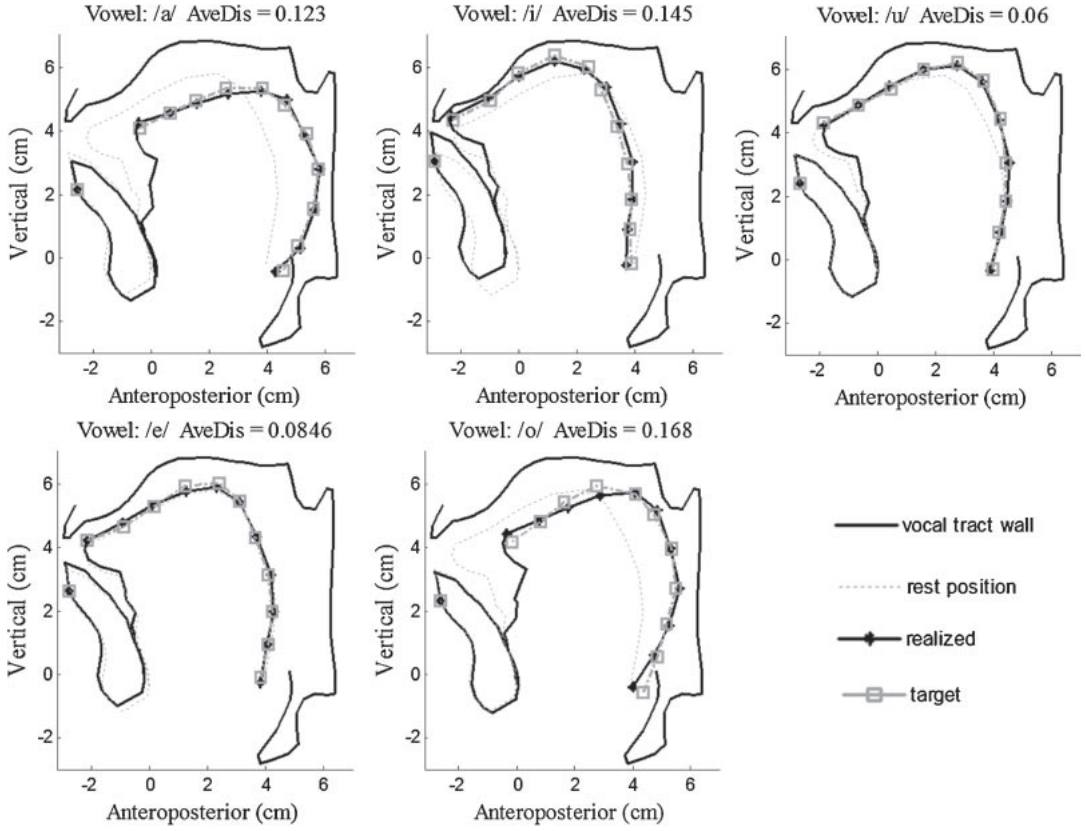

Fig. 14 Realized position for five Japanese vowels. Gray dash lines show the rest position. Gray dash lines with square markers show the target postures. Black lines with stars show the realized positions. The average distances (defined in Eq. (12)) for /a/, /i/, /u/, /e/, and /o/ are 0.123 cm, 0.145 cm, 0.06 cm, 0.085 cm, and 0.168 cm, respectively.

Table 2 Muscle activation patterns for the five Japanese vowels. The active muscle forces of JO and JC are the sum of the active force included in the muscle groups (Unit: Newton).

	GGa	GGm	GGp	HG	SG	SLa	SLp	IL	Va	Vm	Vp	Ta	Tm	Tp	GH	MH	JO	JC
/a/	0	1.78	1.71	6.51	6.04	0	0.11	4.16	1.30	2.67	0	1.76	1.72	1.14	0	0	9.00	0
/i/	0.62	0.51	3.36	0	0	0.02	0	0.78	0	0.41	0	0	0.52	0	0	3.01	0	14.0
/u/	0	0	0	0	3.12	0	0.09	0	0	0	0.67	0	0.87	0	0	0	0	3
/e/	0	0	1.79	0	0	0	0	0.76	0	0	0	0	0.82	0	0	0	0	0
/o/	0	0.48	0	5.13	10.0	0	0.07	0	0	0	0	0	1.68	1.50	0	0	0	0

By the proposed method the target postures of the five Japanese vowels were well achieved. At the same time, we obtained the estimated muscle activation patterns. The target and achieved postures are shown in Fig. 14. The difference, calculated using Eq. (12), ranged from 0.06 to 0.168 cm. The obtained muscle activation patterns (active muscle forces) are shown in Table 2.

To evaluate the obtained muscle activation patterns, we first compare the activations of extrinsic tongue muscles to the normalized EMG (electromyography) measurements [17]. Note that the EMG signals used here were extracted from English vowel articulations because there are no EMG signals for Japanese vowels. The EMG signals and muscle activation were normalized to values between 0 and 1, relative to their maximum value in the activation pattern; the maximum values were normalized to 1. Figure 15 shows the estimated extrinsic tongue muscle activations and the EMG observations. One can see that the estimated muscle activation patterns are consistent with the EMG patterns for vowels /a/, /o/, and /i/. In Fig. 15, there are considerable differences for vowels /e/ and /u/. Japanese /e/ was used for the prototype of the model. Accordingly, there should be no muscle activation in the estimation for vowel /e/. A slight activation of GGp in the estimation was probably caused by the difference in the prototype /e/ and the reference /e/ used in this study. The difference shown for vowel /e/ does not result in any significant effects. As is well known, unlike English /u/, Japanese /u/ does not

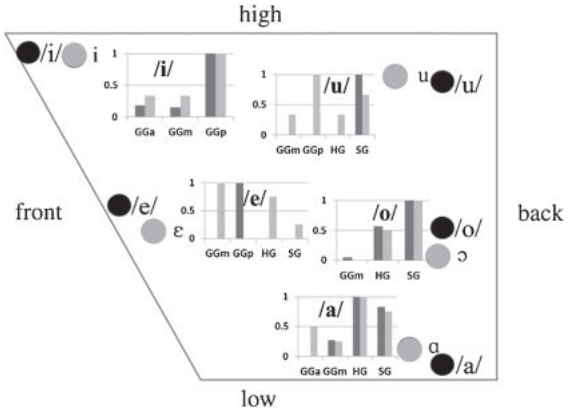

Fig. 15 Estimated extrinsic tongue muscle activations and EMG measurements. The vowel positions are referred from [29]. Black dots represent the positions for Japanese vowels and gray dots show the positions for partial English vowels that are close to the Japanese vowels in articulatory space. Black bars are the normalized results obtained by the proposed method, and the gray bars are the corresponding normalized EMG measurements.

have lip protrusion. The articulatory positions are different between Japanese vowels and English vowels. From this figure, one can see that the vowels used for comparison have similar but not exactly the same articulatory positions. This articulation difference may cause some compensation on the tongue shape, not only on the tongue dorsum. This may be one reason for the difference for /u/.

From Table 2, one can see that intrinsic muscles were activated for all five Japanese vowels, which showed their importance in vowel production. The assessment of the importance of intrinsic muscles in vowel production is very difficult by EMG. The automatic estimation method proposed in this study provides a convenient approach for investigating the functions of intrinsic muscles.

5. DISCUSSION AND CONCLUSION

It is difficult to investigate muscle activations in detail by experimental methods alone. However, a physiological articulatory model can help to overcome this difficulty by predicting muscle activations through simulation. Fang *et al.* have implemented their physiological articulatory model to estimate muscle activities for the five Japanese vowels [4]. In the method reported by Fang *et al.*, the basic process was that the muscle activations were initiated based on EMG observations, and then certain muscles were manually added or removed to reduce the distance between the realized posture and target posture. Their method depends on the expertise of the researcher and manual "trial-and-error" tuning. In order to avoid the deficiency of the manual method, an automatic estimation method was proposed in this study, whereby muscle activations are estimated by systematically and gradually reducing the distance between the realized posture and target posture in accordance with the inherent function of individual muscles. In Fang *et al.*'s method, the full 3D shape of the tongue was used to represent the articulatory posture. In the current study, the midsagittal contour of the tongue and jaw were used to represent the articulatory posture for two reasons: 1) the changes in the midsagittal contour of the tongue and jaw have the greatest effect on the acoustical outcome (especially in the lower range of acoustical frequencies), and 2) using current techniques, it is easier to obtain movement data of the tongue and jaw in the 2D midsagittal plane.

Theoretically, the same articulatory posture may be generated by different muscle activation patterns because muscle activations have more degrees of freedom than does articulatory posture. In order to obtain the optimal activation, economy of energy is typically used as the optimality criterion. Stavness *et al.* [6] proposed a method for finding muscle activations that control the tongue tip to move along given target trajectories by considering minimum muscle activation as a constraint. In this study, although it is difficult to guarantee that the obtained muscle activation has a minimum activation cost, the result can be regarded as a good approximation of the minimum because, in each iteration step, the added muscle gives the greatest contribution to the target vector.

In a muscular-hydrostat system, such as the tongue, muscle orientations change during articulation, which results in a variation of each muscle's function. In this study, a dynamic PCA workspace was constructed to estimate individual muscle functions during articulation. The proposed method was assessed using the articulatory postures of the five Japanese vowels. For vowels far from the "neutral vowel" /e/, the estimated muscle activation patterns are consistent with anatomical knowledge and previously published measurement data. The midsagittal contour including the tongue and jaw was used as the articulatory target, instead of using three crucial points [3,27]. We expect that by using the articulatory posture as a target, the accuracy of model control for speech production will be improved, because the detailed characteristics of speech sounds depend on the whole vocal tract shape rather than the constriction position alone.

In the future, muscle activation patterns for consonants will be investigated by the proposed method and running speech will be generated by using predicted muscle activations based on the physiological articulatory model.

ACKNOWLEDGMENT

The authors would like to thank Dr. Stéphanie Buchaillard for her diligent work on updating the physiological articulatory model to continuum FEM. The authors

thank Professor Kiyoshi Honda, Dr. Hirokazu Tanaka, Dr. Atsuo Suemitsu, and Dr. Shinichi Kawamoto for their helpful comments.

This work is supported in part by the National Basic Research Program of China (No. 2013CB329301), and in part by the National Natural Science Foundation of China (No. 61233009). This study is also supported in part by a Grant-in-Aid for Scientific Research of Japan (No. 25330190) and Grant-in-Aid for Scientific Research (A) (No. 25240026).

REFERENCES

[1] J. Perkell, "A physiological-oriented model of tongue activity in speech production," Ph.D. thesis, MIT (1974).
[2] Y. Payan and P. Perrier, "Synthesis of VCV sequences with a 2D biomechanical tongue shape in vowel production," *Speech Commun.*, **22**, 185–206 (1997).
[3] J. Dang and K. Honda, "Construction and control of a physiological articulatory model," *J. Acoust. Soc. Am.*, **115**, 853–870 (2004).
[4] Q. Fang, S. Fujita, X. Lu, and J. Dang, "A model-based investigation of activations of the tongue muscles in vowel production," *Acoust. Sci. & Tech.*, **30**, 277–287 (2009).
[5] S. Buchaillard, P. Perrier and Y. Payan, "A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning," *J. Acoust. Soc. Am.*, **126**, 2033–2051 (2009).
[6] I. Stavness, J. E. Lloyd and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," *J. Biomech.*, **45**, 2841–2848 (2012).
[7] D. G. Thelen and F. C. Anderson, "Using computed muscle control to generate forward dynamic simulations of human walking from experimental data," *J. Biomech.*, **39**, 1107–1115 (2006).
[8] S. R. Hamner, A. Seth and S. L. Delp, "Muscle contributions to propulsion and support during running," *J. Biomech.*, **43**, 2709–2716 (2010).
[9] S. Sueda, A. Kaufman and D. K. Pai, "Musculotendon simulation for hand animation," in *ACM SIGGRAPH Papers*, pp. 1–8 (2008).
[10] S. Fujita, J. Dang, N. Suzuki and K. Honda, "A computational tongue model and its clinical application," *Oral Sci. Int.*, **4**, 97–109 (2007).
[11] J. Lloyd, I. Stavness and S. Fels, "ArtiSynth: a fast interactive biomechanical modeling toolkit combining multibody and finite element simulation," in *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery* (Springer, Berlin, 2012), pp. 355–394.
[12] T. J. R. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis* (Dover, New York, 2000).
[13] FA. Potra, M. Anitescu, B. Gavrea and J. Trinkle, "A linearly implicit trapezoidal method for integrating stiff multibody dynamics with contact, joints, and friction," *Int. J. Numer. Methods Eng.*, **66**, 1079–1124 (2006).
[14] I. Stavness, J. E. Lloyd, Y. Payan and S. Fels, "Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics," *Int. J. Numer. Methods Biomed. Eng.*, **27**, 367–390 (2011).
[15] J. Dang and K. Honda, "A physiological model of a dynamic vocal tract for speech production," *Acoust. Sci. & Tech.*, **22**, 415–425 (2001).
[16] H. Takemoto, "Morphological analysis of the tongue musculature for three dimensional modeling," *J. Speech Hear. Res.*, **44**, 95–107 (2001).
[17] T. Baer, J. Alfonso and K. Honda, "Electromyography of the tongue muscle during vowels in /əpvp/ environment," *Ann. Bull. RILP*, **7**, 7–18 (1988).
[18] A. Morecki, "Modeling, mechanical description, measurements and control of the selected animal and human body manipulation and locomotion movement," in *Biomechanics of Engineering Modeling, Simulation, Control*, A. Morecki, Ed. (Springer, New York, 1987), pp. 1–28.
[19] V. Hill, "The heat of shortening and the dynamic constants of muscle," *Proc. R. Soc. Lond., Ser. B*, **126**, 136–195 (1938).
[20] R. Wilhelms-Tricarico, "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," *J. Acoust. Soc. Am.*, **97**, 3085–3098 (1995).
[21] R. Laboissière, D. Ostry and A. Feldman, "The control of multimuscle system: Human jaw and hyoid movement," *Biol. Cybern.*, **74**, 373–384 (1996).
[22] V. Sanguineti, R. Laboissière and Y. Payan, "A control model of human tongue movements in speech," *Biol. Cybern.*, **77**, 11–22 (1997).
[23] K. Honda, "Physiological processes of speech production," in *Handbook of Speech Processing* (Springer, Berlin, 2008) pp. 7–26.
[24] M. Stone, "Modeling the motion of the internal tongue from tagged cine-MRI images," *J. Acoust. Soc. Am.*, **109**, 2974–2982 (2001).
[25] S. Takano and K. Honda, "An MRI analysis of the extrinsic tongue muscles during vowel production," *Speech Commun.*, **49**, 49–58 (2007).
[26] K. Miyawaki, "A study of the musculature of the human tongue," *Ann. Bull. Res. Inst. Logop. Phoniatr. Univ. Tokyo*, **8**, 23–50 (1974).
[27] J. Dang and K. Honda, "Estimation of vocal tract shapes from speech sounds with a physiological articulatory model," *J. Phonet.*, **30**, 511–532 (2002).
[28] Q. Fang and J. Dang, *Physiological Articulatory Model for Investigating Speech Production Modeling and Control* (VDM Verlag Dr. Müller, Saarbrücken, 2009), pp. 117–118.
[29] P. Ladefoged and K. Johnson, *A Course in Phonetics*, (Cengage Learning, Boston, 2011).