Biol. Cybern. 91, 275–282 (2004)
DOI 10.1007/s00422-004-0510-6

Biological Cybernetics

Dynamical simulation of speech cooperative articulation by muscle linkages

Takayuki Ito[1], **Hiroaki Gomi**[1,2], **Masaaki Honda**[3]

[1] NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198, Japan
[2] Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuda-cho, Midori-ku, Yokohama, Kanagawa, 226-8503, Japan
[3] Waseda university, 2-579-15 Mikajima, Tokorozawa, Saitama, 359-1192, Japan

Received: 4 July 2003 / Accepted: 15 July 2004 / Published online: 22 September 2004

Abstract. Different kinds of articulators, such as the upper and lower lips, jaw, and tongue, are precisely coordinated in speech production. Based on a perturbation study of the production of a fricative consonant using the upper and lower lips, it has been suggested that increasing the stiffness in the muscle linkage between the upper lip and jaw is beneficial for maintaining the constriction area between the lips (Gomi et al. 2002). This hypothesis is crucial for examining the mechanism of speech motor control, that is, whether mechanical impedance is controlled for the speech motor coordination. To test this hypothesis, in the current study we performed a dynamical simulation of lip compensatory movements based on a muscle linkage model and then evaluated the performance of compensatory movements. The temporal pattern of stiffness of muscle linkage was obtained from the electromyogram (EMG) of the orbicularis oris superior (OOS) muscle by using the temporal transformation (second-order dynamics with time delay) from EMG to stiffness, whose parameters were experimentally determined. The dynamical simulation using stiffness estimated from empirical EMG successfully reproduced the temporal profile of the upper lip compensatory articulations. Moreover, the estimated stiffness variation significantly contributed to reproduce a functional modulation of the compensatory response. This result supports the idea that the mechanical impedance highly contributes to organizing coordination among the lips and jaw. The motor command would be programmed not only to generate movement in each articulator but also to regulate mechanical impedance among articulators for robust coordination of speech motor control.

1 Introduction

Speech articulations, which produce sound pressure variations that are perceived by a listener as linguistic information, are readily accomplished by skillful cooperative movements (Stevens 1998). Investigating this cooperative mechanism is imperative not only for constructing speech models but also for illuminating the brain function involved in speech production. Previous research (Abbs et al. 1984; Folkins and Abbs 1975; Gracco and Abbs 1985; Kelso et al. 1984; Shaiman 1989) has investigated compensatory articulatory movements of the upper lip induced by perturbation of the jaw or lower lip in the production of the bilabial explosive consonants /p/ or /b/. The compensatory movements act effectively to achieve the intended acoustic sounds against unpredictable perturbation. Because the corresponding EMG activity of the primary upper-lip muscle (OOS) increased, it was suggested that an active compensatory mechanism, recruited by somatosensory feedback, contributes to the generation of these compensatory movements. However, the time delay due to nerve conduction and mechanochemical dynamics might present a problem for explaining the rapid regulation of fast speech movements by sensorimotor coordination. Actually, in a recent study of jaw perturbation (Gomi et al. 2002), when the jaw was forced to open by perturbation during the production of the bilabial fricative consonant /Φ/, we found that the upper-lip compensatory response, which contributed effectively to maintaining a labial aperture, was faster than the EMG response of the OOS. This suggested that the compensatory movement is driven not by sensory feedback, but by another, more rapidly acting mechanism.

In a mechanical system, regulating a mechanical impedance, such as stiffness, could induce a rapid reaction against a disturbance. Arm studies have shown that a considerably high mechanical impedance is required to stabilize posture in interaction with the external environment (Hogan 1984; Mussa-Ivaldi et al. 1985) and that stiffness spatial characteristics (orientation and shape of stiffness ellipse) at the hand position can be changed by altering the cocontraction ratio of arm joints (Gomi and Osu 1998). Based on these findings, we hypothesized an alternative mechanism where preprogrammed stiffness of the muscle linkage between the articulators induces the lip compensatory movements in order to maintain the fricative sound. To examine this hypothesis,

Correspondence to: T. Ito
(e-mail: taka@idea.brl.ntt.co.jp,
Tel.: +81-46-2403592, Fax: +81-46-2404721)

Gomi et al. (2002) proposed a method of stiffness estimation in which a relative stiffness value is calculated for the perturbed displacements. This method enabled us to compare the stiffness between different timings in speech. The results indicated that muscle stiffness varied according to the speech task and there was a large compensatory movement of the upper lip under the high-stiffness condition. However, the dynamical forces of inertia and viscosity were not taken into account in estimating stiffness, mainly because of experimental limitations. If these forces were great, the estimated stiffness would be imprecise.

To verify the hypothesis of Gomi et al. (2002) that initial compensatory movements are due to preprogrammed muscle stiffness, we reproduced the upper-lip compensatory movement by using the upper-lip jaw dynamical model, in which the stiffness is varied according to muscle activity level. In Sect. 2, the complicated muscle connections between the upper lip and jaw are modeled by a simple mechanical dynamics (inertia, viscosity, and stiffness). In Sect. 3, in order to estimate the stiffness variation from the muscle activation level, the transformation from EMG signal to muscle stiffness is modeled as second-order dynamics with time delay and identified from independent experimental data that were derived from responses to electrical stimulation. We examine the temporal pattern of the reproduced lip movement and the modulation of compensatory articulations according to the particular speech task (/Φ/ utterance) by comparing simulated behaviors with corresponding actual behaviors.

2 Modeling the upper-lip dynamics

The mechanical linkage model of the upper lip–jaw is shown in Fig. 1a. The jaw (mass: m_j), upper lip (mass: m_u), and upper lip elevating muscles, including soft tissues, are serially connected by two springs (stiffness: k_1, k_2) and two dampers (viscosity: c_1, c_2). Based on the facts that muscles have viscoelasticity (Gomi and Osu 1996; Kearney and Hunter 1990; Osu and Gomi 1999) and that characteristics have often been modeled as a linear spring (with damper) (Agarwal and Gottlieb 1977; Kearney and Hunter 1982; Bennett et al. 1992; Lacquaniti et al. 1993), the upper-lip elevation muscles (ULEM), which consist of levator labii superior, levator anguli oris, zyogomaticus major, and zyogomaticus minor, are modeled as the upper spring k_1 and damper c_1, and the muscles between the upper lip and jaw [OOS and depressor angri oris (DAO)] are modeled as the lower ones. According to this model, the upper-lip dynamics can be represented by the following equation:

$$\begin{aligned} & m_u\ddot{x}_u + c_1\dot{x}_u - c_2(\dot{x}_j - \dot{x}_u) \\ & = -k_1(x_u^e - x_u) + k_2\left\{(x_j^e - x_j) - (x_u^e - x_u)\right\}, \end{aligned} \tag{1}$$

where x_u^e and x_j^e are the equilibrium trajectories, which represent the time series of equilibrium positions of muscle dynamics. The first term on the right-hand side indicates the force generated by the upper spring and the second term that generated by the lower spring (Fig. 1a).

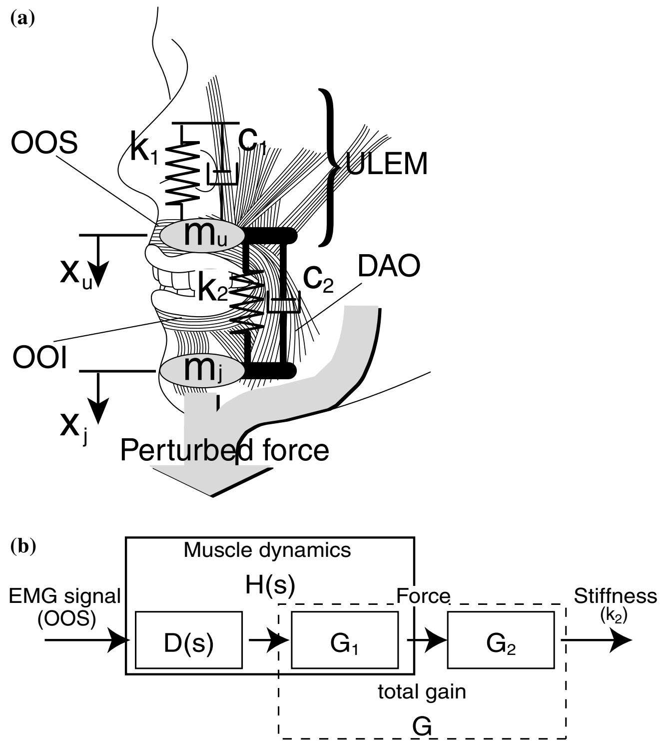

Fig. 1. **a** Model of mechanical linkage of upper lip–jaw. **b** Stiffness estimation from EMG signal

The viscosity coefficients, c_1 and c_2, are set in proportion to the stiffness k_1 and k_2, respectively, as follows:

$$c_i = \alpha k_i \quad (i = 1, 2)\,. \tag{2}$$

Previous studies of limb biomechanics have shown a roughly linear relationship between muscle stiffness and EMG activity under isometric conditions (Gomi and Osu 1998; Kearney and Hunter 1990). Other studies have shown that the joint stiffness (or force) can be estimated from an EMG signal (Osu and Gomi 1999; Meek et al. 1990; Koike and Kawato 1995). Thus, for the articulatory organs it is quite likely that stiffness of the muscle linkage between articulators varies according to the muscle activity. Gomi et al. (2002) demonstrated that there is no remarkable change in EMG activity of ULEM during a sentence and that the EMG activity of the OOS and DAO increases for the production of the bilabial fricative consonant /Φ/. These observations suggest that stiffness k_1 does not vary and stiffness k_2 varies according to the speech task. Thus, we define k_1 as being constant and k_2 as being dependent on the EMG activity level.

3 Muscle dynamics of the upper lip

3.1 Modeling

Obtaining the stiffness k_2 from an EMG signal is based on the following two facts. The first is that the transformation between the EMG and force, $H(s)$, can be represented as second-order dynamics with a time delay, $D(s)$, such that

$$H(s) = G_1 D(s) = G_1 \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + {\omega_n}^2} e^{-s\tau}, \quad (3)$$

where ω_n denotes the natural frequency, ζ the damping ratio, τ the time delay, G_1 the gain, and s a Laplacian operator, as studied in skeletal muscle modeling (Mannard and Stein 1973; Koike and Kawato 1995). The temporal variation of the force estimated from an EMG signal was determined according to the dynamics $D(s)$. This relationship is referred to here as muscle dynamics. The second relationship, according to previous studies (Kearney and Hunter 1990; Gomi and Osu 1998) that describe the linear relationship between stiffness and torque, is that the estimated force is linearly transformed into stiffness by a proportional gain G_2. This transformation from EMG signal to stiffness is illustrated in the block diagram in Fig. 1b. In the next section we will describe how the dynamic part $D(s)$ (ω_n, ζ, and τ) is estimated. Gains G_1 and G_2 are determined together as the total gain ($G = G_1 \times G_2$) in Sect. 4.2.

3.2 Experiment and parameter identification

To obtain the muscle activation impulse response, we induced a low-level contraction of the lip muscle by electrically stimulating the motor nerves that innervate the OOS and measured the upper-lip force generated by the contraction. Three subjects [two male (A and B), one female (C)], who had never experienced peripheral neuropathy of the facial nerves, participated in this experiment. The subject's upper lip rested on a cantilever beam, whose opposite end was attached to a six-axis force sensor (Nitta UFS-3012A15). The force signal was sampled at 2 kHz. The pulse stimulus signal (duration: 300 μs) was generated by an electrical stimulator (Nihon Kohden SEN-3301) and isolator (Nihon Kohden SS-104J) and excited the motor nerve that innervates the OOS with a pair of surface electrodes (Nihon Kohden NM-430S). The stimulus site with the largest upper-lip response was determined by searching exhaustively beneath the zygomatic bone, under which lie the motor nerves that innervate the OOS. We confirmed by visual inspection or EMG (ipsilateral OOS, DAO, and ULEM) that other facial muscles, such as orbicularis oculi, were hardly stimulated. The stimulus intensity was set at a painless level for each subject.

The obtained muscle force response (average of 100 trials) is represented as the solid line in Fig. 2. This response can be regarded as the impulse response of the muscle dynamics in (3) since the pulse signal was used as an input stimulus. The best-fit parameter values [ω_n, ζ and τ of $D(s)$] were obtained by nonlinear optimization using the "lsqcurvefit" function in the MATLAB software (The Mathworks, Inc.). The identified parameter values of ω_n, ζ, and τ in (3) are shown in Table 1. Note that G_1 is the gain for the electrical stimulation (not for actual neural input, because of the impedance of the skin and other orofacial tissue), so it was not evaluated here. How we determined G_1 as used in the simulation is described in Sect. 4.2. The dotted line in Fig. 2 shows the best-fit impulse response corresponding to the experimental data [subject A, variance accounted for (VAF) = 0.99]. The mean and standard deviation of VAF for all subjects was 0.99 ± 0.001.

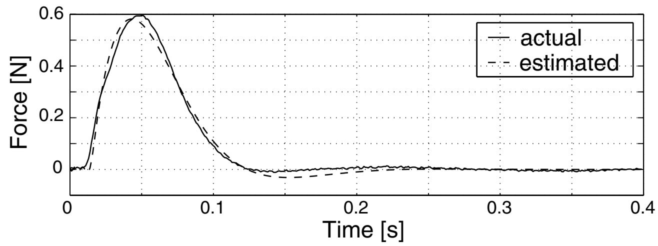

Fig. 2. Measured (*solid*) and estimated (*dotted*) force responses of the upper lip elicited by electrical stimulation of the motor nerve (subject A). The *left end* of the horizontal axis indicates the start time of the stimulus. The measured force signal is the average of 100 trials. The estimated force response is derived from the impulse response of muscle dynamics estimated by nonlinear optimization

Table 1. Muscle dynamics parameters estimated by nonlinear optimization

	Subject			
	A	B	C	Average
Natural frequency (ω_n) (Hz)	6.26	5.54	5.29	5.70 (0.50)
Damping ratio (ζ)	0.682	0.740	0.842	0.755 (0.08)
Time delay (τ) (ms)	14.5	15.7	12.3	14.2 (1.72)
RMS error/Peak response (%)	3.17	3.53	3.54	3.41 (0.21)

4 Dynamical simulation

4.1 Jaw perturbation experiment

In previous studies (Gomi et al. 2002; Ito et al. 2000), we observed the compensatory movement of the upper lip in response to jaw perturbation. Details of the system, experimental procedure, and data collection are described elsewhere (Gomi et al. 2002). In brief, four subjects (all males, Japanese native speakers: A, B, D, and E) were instructed to say a carrier sentence "kono /$a\Phi a\Phi a$/ mitai" for each trial. The subject's jaw was connected to a jaw manipulandum system, which can disturb or assist jaw movement. Note that this connection did not interfere very much with speech movements under the unperturbed condition. A downward perturbation, which was stepwise in shape (4.0 N), disturbed the jaw movement by suddenly acting in the jaw-open direction and was triggered 0 (P_1), 30 (P_2), 60 (P_3), 90 (P_4), or 120 (P_5) ms (the five triangles in the bottom graph of Fig. 3) after the start time of jaw elevation during the /a_1/ utterance. Fifty trials were randomly selected among all trials (500) as perturbed trials. The EMG signal of the OOS was amplified and filtered (bandpass: 50–1,500 Hz) with biomedical-amplifier (Nihon Kohden, MME-3116) and sampled at 2 kHz. Typical data are shown in the middle panel in Fig. 3. EMG activity of the OOS increased

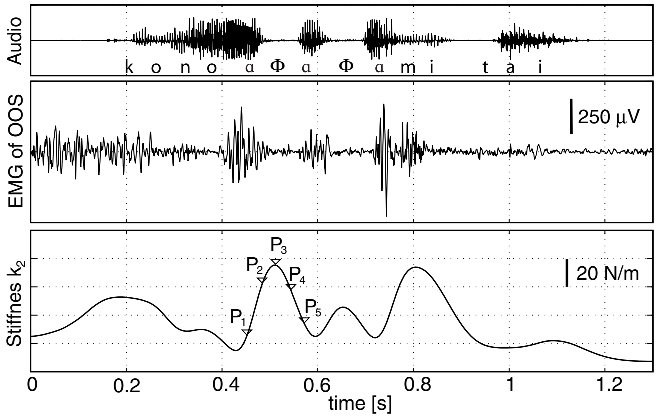

Fig. 3. Measured audio and raw EMG data of OOS are shown in *top* and *middle panels*, respectively (one typical trial, subject A). The *bottom panel* shows a temporal pattern of stiffness $k_2[= H(s)E(t)]$ in the simulation. $E(t)$ was obtained by rectifying, smoothing, and averaging raw EMG data. The *five triangles* indicate perturbation onsets ($P_1 \sim P_5$)

prior to the production of labial utterances (/Φ/and/m/) because of a delay of muscle contraction and the influence of inertia. Articulatory movements [upper lip (UL), lower lip (LL + J: lower lip position in space), and jaw (J)] were measured at 250 Hz with a three-dimensional optical position sensor (OPTOTRAK 3020). In Fig. 4a, the dotted line indicates perturbed trajectory, which was obtained by adding the perturbation-induced component to the control trajectory. The LL + J was suddenly shifted by jaw perturbation and almost recovered the control trajectory around production of /Φ_2/. The upper lip immediately shifted downward in response to the jaw perturbation. Consequently, this upper-lip shift compensated for labial distance widened by jaw perturbation to produce /Φ_1/ utterance. In addition, for /*a*/ utterance, which does not require keeping a precise labial constriction, the response was less induced than that for /Φ/ utterance. Thus, this compensatory response was functionally modulated according to task requirements.

4.2 Simulation procedure

As shown in Fig. 1a, the jaw-perturbing force comprises two components. One force component affects the upper lip–jaw spring k_2 in our model, and the other, the head–jaw spring, which consists of jaw-closing muscles, such as the masseter and tempolaris (not described in Fig. 1a). Since the jaw inertia and the stiffness of the jaw-closing muscles are not known, the jaw movement cannot be predicted by a dynamical equation; consequently, the interaction force between the upper lip and jaw cannot be estimated in this way. To estimate it, we utilized the experimentally obtained perturbed jaw trajectory instead. The perturbed jaw trajectory x_j^p can be expressed by the sum of its control trajectory x_j^c and the perturbation-induced component $\delta x_\mathrm{j}^\mathrm{p}$ as

$$x_\mathrm{j}^\mathrm{p} = x_\mathrm{j}^\mathrm{c} + \delta x_\mathrm{j}^\mathrm{p}.$$

By replacing x_j in (1) by x_j^p,

$$\begin{aligned} & m_\mathrm{u}\ddot{x}_\mathrm{u} + c_1\dot{x}_\mathrm{u} - c_2(\dot{x}_\mathrm{j} - \dot{x}_\mathrm{u}) \\ &= -(k_1+k_2)\left(x_\mathrm{u}^\mathrm{e} - x_\mathrm{u}\right) + k_2\left\{x_\mathrm{j}^\mathrm{e} - (x_\mathrm{j}^\mathrm{c} + \delta x_\mathrm{j}^\mathrm{p})\right\} \\ &= -(k_1+k_2)\left\{x_\mathrm{u}^\mathrm{e} - x_\mathrm{u} - \frac{k_2\left(x_\mathrm{j}^\mathrm{e} - x_\mathrm{j}^\mathrm{c}\right)}{(k_1+k_2)}\right\} - k_2\delta x_\mathrm{j}^\mathrm{p} \\ &= -(k_1+k_2)(v - x_\mathrm{u}) - k_2\delta x_\mathrm{j}^\mathrm{p}, \end{aligned} \tag{4}$$

where

$$v = x_\mathrm{u}^\mathrm{e} - \frac{k_2(x_\mathrm{j}^\mathrm{e} - x_\mathrm{j}^\mathrm{c})}{k_1+k_2}.$$

Here, the first term on the right side of the last line of (4) is the force applied to the upper lip by both springs under an unperturbed (control) condition, and the second term corresponds to the perturbation-induced forces. The upper-lip behavior perturbed at the particular timing ($P_1 \sim P_5$) could be simulated by applying the perturbation-induced trajectory $\delta x_\mathrm{j}^\mathrm{p}$ of the corresponding timing. In order to simulate the upper-lip movement x_u, the corresponding trajectory v has to be determined. In this case, v can be obtained by using the control trajectories ($\ddot{x}_\mathrm{u}^\mathrm{c}$, $\dot{x}_\mathrm{u}^\mathrm{c}$, x_u^c, and $\dot{x}_\mathrm{j}^\mathrm{c}$) with $\delta x_\mathrm{j}^\mathrm{p} = 0$ (for all temporal steps) in (4) as follows:

$$v = \frac{-m_\mathrm{u}\ddot{x}_\mathrm{u}^\mathrm{c} - c_1\dot{x}_\mathrm{u}^\mathrm{c} + c_2(\dot{x}_\mathrm{j}^\mathrm{c} - \dot{x}_\mathrm{u}^\mathrm{c})}{k_1+k_2} + x_\mathrm{u}^\mathrm{c}. \tag{5}$$

Thus, by using the trajectory v, control trajectory x_u^c can be reproduced in dynamical simulation with $\delta x_\mathrm{j}^\mathrm{p} = 0$.

The control trajectories x_u^c and x_j^c were obtained by averaging $70 \sim 100$ out of 450 control trajectories. The control trials were selected so that there would be a small deviation from each perturbed trajectory during the period between the speech sound onset and perturbation onset. The perturbation-induced component, $\delta x_\mathrm{j}^\mathrm{p}$, was obtained for each perturbation pattern ($P_1 \sim P_5$) by subtracting the control trajectory, x_j^c, from perturbed jaw trajectories x_j^p [see Gomi et al. (2002) for details]. The jaw velocity $\dot{x}_\mathrm{j}$ was obtained as the first derivative of measured jaw trajectory x_j.

The mechanical parameters could not be obtained experimentally because the jaw-perturbing force could not be determined, as described at the beginning of this section. We instead determined them as follows. The mass of the upper lip m_u was set at 0.05 kg as a rough estimation. Considering previous studies of cross-joint arm muscles (0.016–0.031) (Gomi and Osu 1996, 1998), the ratio in (2) was fixed at $\alpha = 0.02$. Both were fixed for all subjects throughout. The temporal variation of the stiffness k_2, i.e., $D(s) \times$ EMG, was calculated from averaged EMG signals during the unperturbed utterance "kono /*a*Φ*a*Φ*a*/ mitai". The EMG patterns for averaging were selected among all the control trials. The total gain $G(= G_1 \times G_2)$ for obtaining stiffness k_2 (Fig. 1b) and stiffness k_1 was determined to minimize the sum of the squared error between the actual and simulated shifts of the upper lip only for two data points [40 ms after the perturbation onsets in two of the five cases ($P_1 \sim P_5$)]. (See Fig. 3, a particular case, as an example.)

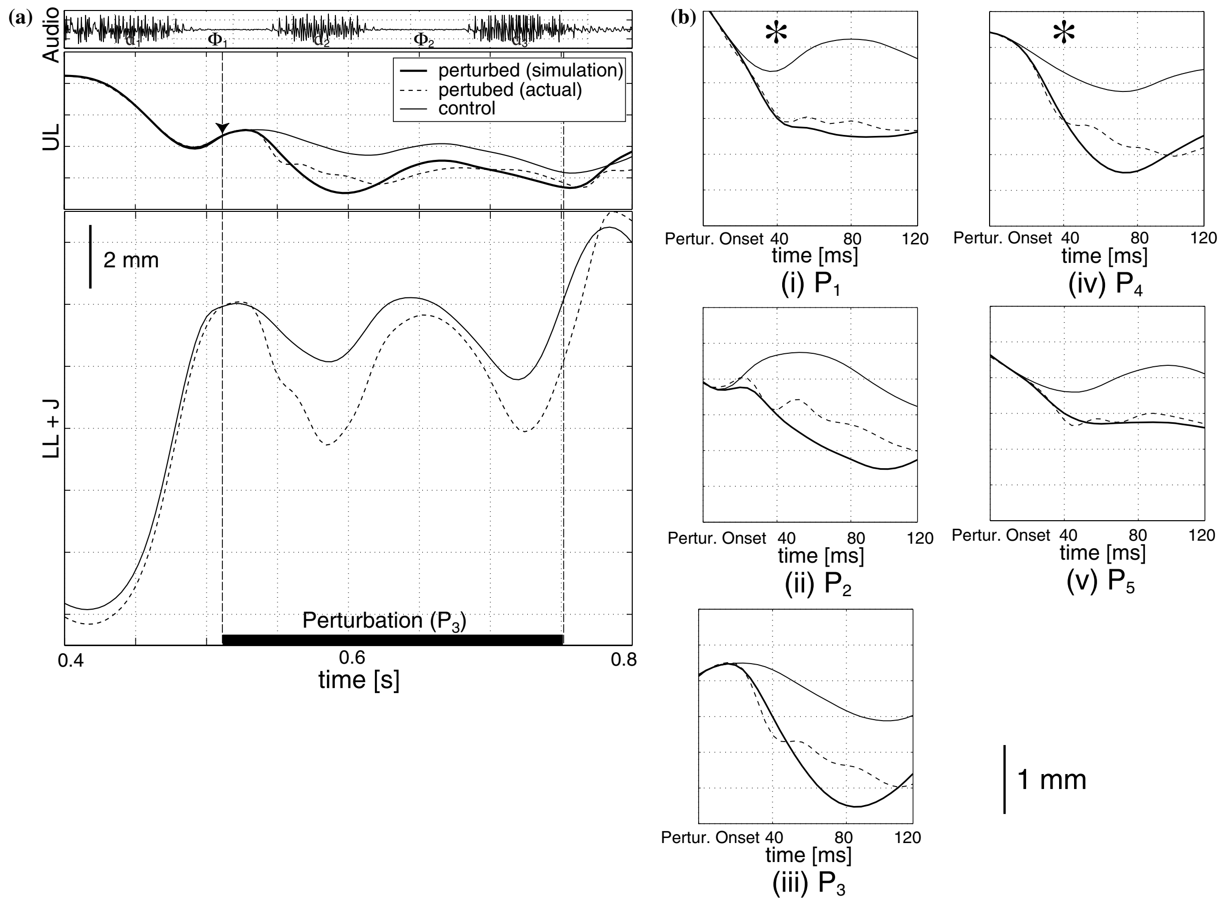

Fig. 4. **a** An audio signal of the utterance /*a*Φ*a*Φ*a*/, trajectories of the upper lip [simulation (*solid thick line*), and experimental (*dotted line*: perturbed, *solid-thin line*: control)] and trajectories of the lower lip for the perturbation P_3 (subject A). The *thick bar* indicates the duration of perturbation. *Vertical dotted lines* indicate perturbation onset of P_3 and release. **b** The upper-lip trajectories for all perturbation timings ($P_1 \sim P_5$) from perturbation onset to 120 ms after perturbation onset (subject A). The two displacements 40 ms after perturbation onset (marked "$*$") were used for gain adjustment of the model

5 Results

5.1 Estimated stiffness

As shown in the bottom panel of Fig. 3, the stiffness k_2 varied during the speech task, which was determined by using muscle dynamics from the EMG signal and optimization of gain G. The three peaks of stiffness k_2 in the range of $0.4 \sim 1.0$ s corresponded to an increase in EMG activity for labial utterances (/Φ/ and /m/). As indicated by the five triangles, stiffness values were different in each perturbation onset ($P_1 \sim P_5$). Since the stiffness, k_1, is constant, this variable stiffness, k_2, would be critical to a response variation for the upper-lip compensation according to the jaw perturbation. The maximum value of the stiffness k_2 was 108 N/m among all subjects. Mean and standard deviations of k_1 were 120 ± 31.1 N/m. Although, as described in Sect. 2, the EMG of ULEM showed a low activity compared with the OOS and DAO during /Φ/ utterance, the k_1 value is relatively higher than the k_2 value. This might be due to passive tight linkage between the perioral region and skull.

5.2 Temporal profile of the upper-lip compensatory articulation

The middle panel of Fig. 4a shows a typical simulated upper-lip trajectory (solid thick line) superimposed on the corresponding perturbed (dotted line) and control (solid thin line) trajectories obtained in the experiment (subject A, P_3). In the initial downward shift ($0 \sim 40$ ms) after perturbation onset the simulated upper-lip trajectory almost mimics the actual perturbed trajectory, indicating that the simulated upper-lip dynamics (i.e., stiffness) is almost the same as the actual one in that phase.

This good performance in simulation is generally consistent for all five perturbation patterns ($P_1 \sim P_5$) as seen in Fig. 4b, which shows the control, actual perturbed, and simulated perturbed trajectories 120 ms after the perturbation onset. In all panels, the downward shifts closely match during the initial 40 ms after perturbation onset. Response variation among perturbations [large and quick responses in (i, iii, iv) and small and slow responses in (ii, v)] are reproduced successfully. Similar results were observed for the other subjects (B, D and E). The correlation coefficient between actual and simulated response variation (displacements 40 ms after perturbation onset in $P_1 \sim P_5$) was 0.83 (±0.09) for the four subjects.

Small discrepancies between actual and simulated responses remain because we used only two data points among 150 [30 points (120-ms, 250-Hz sampling) of five patterns]. The difference at the very beginning of the response (around perturbation onset) to the perturbation observed in Fig. 4b(ii) could be due to an unexpected bias error contained in the averaged EMG pattern of the control trials or in the jaw-perturbed component δx_j^p. After the initial response (i.e., > 40 ms), on the other hand, the simulated trajectory differs from the actual one in P_2, P_3, and P_4 cases [panels (ii–iv), respectively]. These differences could be due to additional influences on the lip movements. Details will be discussed in Sect. 6.2.

5.3 Variable stiffness contribution

Gomi et al. (2002) hypothesized that variable stiffness is important in generating a response variation of the upper-lip compensation according to speech tasks. To evaluate the model performance by muscle stiffness contribution, we calculated the error ratio (ER) between simulated and actual trajectory for initial response (<40 ms):

$$\mathrm{ER} = \sqrt{\frac{\sum_i \left(\delta x_{\mathrm{sim}}^{\mathrm{p}}(P_i) - \delta x_{\mathrm{act}}^{\mathrm{p}}(P_i)\right)^2}{\sum_i \delta x_{\mathrm{act}}^{\mathrm{p}}(P_i)^2}}, \tag{6}$$

where $\delta x_{\mathrm{sim}}^{\mathrm{p}}$ represents the perturbation-induced displacement of ten data points (during a 40-ms period after perturbation onset) × 3 cases (nontraining data sets in $P_1 \sim P_5$) in the simulation, and $\delta x_{\mathrm{act}}^{\mathrm{p}}$ that in the experiment. The mean (±SD) of ER for four subjects was 16.14 (±5.88)%, indicating that actual perturbed lip movement can be mostly represented by the simulation.

To consider the alternative possibility that the stiffness change does not produce the variation of positional shifts, we performed a simulation assuming constant stiffness during the utterance. The stiffness k_2 was assumed to be a constant value × total gain G. The stiffness k_1 and the total gain G were optimized in the same manner as in the variable stiffness case. The other parameters [m_u in (4) and α in (2)] and trajectories [$\dot{x}_j$, x_j^p in (4) and control trajectories in (5)] were the same as those in the variable stiffness simulation. In this case, the simulated upper-lip responses were found to be roughly proportional to the perturbed jaw behaviors. As a result, these simulated responses poorly fitted the actual lip responses to the five perturbations ($P_1 \sim P_5$). The ER = 28.77 (±8.50)% of this case is significantly larger ($p < 0.05$) than that using variable stiffness, indicating that the stiffness variation is required in order to induce the response variability according to speech task.

Moreover, to examine the fitting performance of the muscle dynamics shown in Sect. 3, we also calculated the performance of the simulated upper-lip movements using the alternative muscle model of the finger [Akazawa et al. (1988), natural frequency: 1.73 Hz, damping ratio: 1.29, no time delay]. Note that the total gain G and the stiffness k_1 were adjusted for this simulation. ER = 27.31 (±6.97)% for this model was significantly greater ($p < 0.05$) than for the current muscle model, indicating that the transformation from EMG signal to the muscle stiffness identified in this study is more suitable for describing the articulatory muscle dynamics.

6 Discussion

6.1 Muscle dynamics of force generation

Several studies have attempted to elucidate the relationship between muscle force and EMG signals and concluded that second-order dynamics with a time delay could satisfactorily represent that relationship in humans (finger, Akazawa et al. 1988; arm, Koike and Kawato 1995; jaw, Cooker et al. 1980) and in cat (limb, Mannard and Stein 1973). The identified muscle dynamics in this study has a higher natural frequency (5.7 Hz) than that of the human finger (Akazawa et al. 1988: 1.73 Hz) and jaw (Cooker et al. 1980: 3 Hz), and is similar to the cat limb muscle of Mannard and Stein (1973) (5 Hz). The reason for this difference may be as follows. When we had four subjects attempt to rapidly generate muscle force by voluntary contraction under isometric conditions as in previous studies (Akazawa et al. 1988; Cooker et al. 1980; Koike and Kawato 1995), it was difficult for them to generate OOS force regularly beyond a rate of 4 Hz (unpublished data). Cooker et al. (1980) reported that 4 Hz is the rate limit for the voluntary contraction of the jaw-closing muscles. On the other hand, the motion frequency for the lips during repetitive speech production has been measured as being around 6 Hz in Kelso et al. (1985), and the minimum time taken to maneuver the lips from an unrounded to a rounded configuration is in the range of 50–100 ms, which corresponds to the range of 5–10 Hz (Stevens 1998). These findings suggest that the dynamics relating force to the EMG for lip muscle should have a higher natural frequency than those identified in previous studies. Therefore, EMG and force data from voluntary contraction might be inadequate for identifying muscle dynamics for speech articulation.

In the case of electrical stimulation, especially pulse stimulation, on the other hand, we could readily induce a muscle contraction at a sufficient frequency range. Consequently, we could identify the muscle dynamics with a higher natural frequency as in Mannard and Stein (1973). These data suggest that pulsatile electrical stimulation might be useful in obtaining the physiological limit of muscle response.

6.2 Modeling and simulation performance

In order to examine the task-dependent contribution of the stiffness between the upper lip and jaw, it is very important to determine quantitatively the temporal variation of the stiffness. We here estimated the temporal variation of stiffness from EMG signal based on the fact that the stiffness and EMG are correlated (Agarwal and Gottlieb 1977;

Osu and Gomi 1999). This method allows us to estimate temporal variation of the stiffness. As a result, it enables dynamical simulation of the lip–jaw linkage to explore the coordinative articulatory mechanism, which is under the control of the central nerves system (CNS).

In order to determine the stiffness, the gain parameter $G(=G_1 \times G_2)$ is still needed. However, this could not be measured directly, as mentioned before. Alternatively, this coefficient was determined by minimizing the error of the two data points selected from two of five perturbed behaviors. Thus, the model must predict the other three temporal patterns of the lip movement [in Fig. 4b(ii, iii, v)], which were not used in gain optimization. For these three patterns, the error, ER, in the simulation using variable stiffness was significantly smaller than that using the constant stiffness as described in Sect. 5.3. Note that we also confirmed, by considering muscle size, that the stiffness in the simulation was of the same order of the arm muscle stiffness measured in previous studies (Gomi and Osu 1996; Osu and Gomi 1999). From the above consideration, it is suggested that the temporal stiffness variation is effective in producing the task-dependent cooperative speech articulation.

As described in Sect. 5.2, there were discrepancies between the simulated and actual trajectory after the initial response (>40 ms), especially in $P_2 \sim P_4$ cases. One major cause of these discrepancies could be the upper-lip lateral contact with the lower lip (the contact at the lateral part of the lips near the corner of the mouth). Since the upper-lip downward force was produced around the first upper peak of the upper-lip trajectory, the equilibrium position of the upper lip would move downward. Due to the application of perturbations, the equilibrium position of the upper lip further shifted downward, which would produce large downward shift of the actual position of the upper lip. However, this lowering could be interrupted by the lateral contact with the lower lip. The discrepancy between the actual and simulated trajectories can be ascribed to this interaction effect, which was not included in the simulation dynamics. For the responses of P_1 and P_5, on the other hand, equilibrium positions of the upper lip would not shift downward significantly because of the upward movement of the control trajectory of the upper lip [see panels (i, v) in Fig. 4b]. In addition, neither this contribution of the contact dynamics nor the sensory feedback was taken into account in the current simulation, both of which may affect the upper lip movement over 40 ms after perturbation onset. By introducing these effects, the latter part of the response could be reproduced more accurately, but this requires further discussion and is beyond the main scope of this study.

6.3 Speech articulation target

Due to the complex process of speech production (motor commands, articulator movements, vocal tract configuration and vocal fold vibration, and speech sounds), the target or goal in speech motor execution is controversial. Houde and Jordan (1998) provided evidence of an auditory target in speech articulation by demonstrating formant frequency tracking in auditory alteration. However, Tremblay et al. (2003) recently showed articulatory readjustment in a force field jaw movement without acoustical feedback. This suggests that speech articulation and associated somatosensory inputs constitute a target of speech production (i.e., motor target). The muscle stiffness variation for bilabial fricative production we examined here can be regarded as one control strategy for maintaining the vocal tract configuration. It is therefore important to examine this stiffness hypothesis in speech articulation, which supports the target representation in the "motor" rather than "acoustic" domain.

Because of inherent somatosensory feedback in real human articulation, it would be intriguing to examine the possibility of this stiffness hypothesis in dynamical simulation rather than in experiment. In this study, the initial and dominant components of compensatory articulation (<40 ms) were successfully reproduced by using a mechanical linkage model of the upper lip and jaw. The stiffness transformed from the empirical EMG data nicely reproduced the variation of upper-lip compensatory movements. This suggests that muscle mechanical dynamics plays a significant role in coordinating multiple articulators for achieving speech tasks and could be interpreted to mean that the mechanical impedance, as well as movement, is taken into account in motor control of the orofacial system. Considering the impedance (or stiffness) control of arm studies (Hogan 1984; Mussa-Ivaldi et al. 1985; Gomi and Osu 1998; Burdet et al. 2001; Osu et al. 2001) and the task-dynamics model (Saltzman 1986), it is likely that the CNS regulates motor commands not only for movement but also for dynamical interaction between articulators. In the production of a bilabial fricative consonant, since the increase of muscle stiffness would be required to produce a labial constriction at any jaw position and maintain a precise constriction against airflow passing between the upper and lower lips, the muscle stiffness could be considered a control variable rather than a by-product of movement. This idea is further supported by the observation of Ito et al. (2003) that OOS muscle activity does not correlate with the jaw-closing position for /Φ/ production in unperturbed (normal) speech, while the upper lip position does. That OOS muscle activity is less correlated with upper-lip position might be closely associated with the muscle stiffness regulation. If so, then muscle stiffness would also be represented as the articulatory target.

Note that, even though this study provides supporting evidence for the importance of mechanical dynamics, we do not deny the importance of sensory feedback and auditory feedback mechanisms in speech production. It is evident that speech production is severely impaired by anesthesia and delayed auditory feedbacks (Abbs et al. 1976; Hain et al. 2001). It is necessary to clarify the roles of these feedbacks and mechanical dynamics to understand the precise mechanisms of speech motor control.

Acknowledgements. This research was supported by Core Research for Evaluation Science and Technology (CREST) of the Japan Science and Technology Agency. We thank J. Perkell

(Massachusetts Institute of Technology) for giving us a number of helpful suggestions and improving the manuscript. We also thank T. Hirahara (ATR Human Information Science Laboratories) for his continuing encouragement, E.Z. Murano (ATR Human Information Science Laboratories) for her assistance in the experiment, and T. Konno and M. Sawada (NTT-AT) for their support in software development.

References

Abbs JH, Folkins JW, Sivarajan M (1976) Motor impairment following blockade of the infraorbital: nerve implications for the use of anesthetization techniques in speech research. J Speech Hear Res 19(1):19–35

Abbs JH, Gracco VL, Cole KJ (1984) Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. J Mot Behav 16(2):195–231

Agarwal GC, Gottlieb CL (1977) Compliance of the human ankle joint. Trans ASME 99:166–170

Akazawa K, Takizawa H, Hayashi Y, Fujii K (1988) Development of control system and myoelectric signal processor for bio-mimetic prosthetic hand. Biomechanism 9:43–53

Bennett DJ, Hollerbach JM, Xu Y, Hunter IW (1992) Time-varying stiffness of human elbow joint during cyclic voluntary movement. Exp Brain Res 88:433–442

Burdet E, Osu R, Franklin DW, Milner TE, Kawato M (2001) The central nervous system stabilizes unstable dynamics by learning optimal impedance. Nature 414:446–449

Cooker HS, Larson CR, Luschei ES (1980) Evidence that the human jaw stretch reflex increases the resistance of the mandible to small displacement. J Physiol 308:61–78

Folkins JW, Abbs JH (1975) Lip and jaw motor control during speech: responses to resistive loading of the jaw. J Speech Hear Res 18:207–220

Gomi H, Ito T, Murano EZ, Honda M (2002) Compensatory articulation during bilabial fricative production by regulating muscle stiffness. J Phon 30(3):261–279

Gomi H, Osu R (1996) Human arm stiffness and viscosity in interaction with environments on a horizontal plane. Technical Report ISRL-96-3, NTT Basic Research Laboratories, Information Science Research Laboratory, Kanagawa, Japan

Gomi H, Osu R (1998) Task-dependent viscoelasticity of human multijoint arm and its spatial characteristics for interaction with environments. J Neurosci 18(21):8965–8978

Gracco VL, Abbs JH (1985) Dynamic control of perioral system during speech: kinematic analysis of autogenic and nonautogenic sensorimotor processes. J Neurophysiol 54(2):418–432

Hain TC, Burnett TA, Larson CR, Kiran S (2001) Effects of delayed auditory feedback (daf) on the pitch-shift reflex. J Acoust Soc Am 109(5 Pt 1):2146–2152

Hogan N (1984) Adaptive control of mechanical impedance by coactivation of antagonist muscles. IEEE Trans Auto Cont AC-29(8):681–690

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279(5354):1213–1216

Ito T, Gomi H, Honda M (2000) Task dependent jaw-lip coordination examined by jaw perturbation during bilabial-consonant utterances. In: Proceedings of the 5th seminar on speech production (SPS5), 1–4 May 2000, Kloster Seeon, Bavaria, Germany, pp 41–44

Ito T, Gomi H, Honda M (2003) Articulatory coordination using mechanical linkage between upper lip and jaw examined by jaw perturbation. Trans IEICE J86-DII(2):333–341

Kearney RE, Hunter IW (1982) Dynamics of human ankle stiffness: variation with displacement amplitude. J Biomech 15(10):753–756

Kearney RE, Hunter IW (1990) System identification of human joint dynamics. Crit Rev Biomed Eng 18:55–87

Kelso JAS, Tuller B, Vatikiotis-Bateson E, Fowler CA (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. J Exp Psychol Hum Percept Perform 10(6): 812–832

Kelso JAS, Vatikiotis-Bateson E, Saltzman EL, Kay B (1985) A qualitative dynamics analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. J Acoust Soc Am 77(1):266–280

Koike Y, Kawato M (1995) Estimation of dynamic joint torque and trajectory formation from surface electromyography signals using a neural network model. Biol Cybern 73:291–300

Lacquaniti F, Carrozzo M, Borghese NA (1993) Time-varying mechanical behavior of multijointed arm in man. J Neurophysiol 69(5):1443–1464

Mannard A, Stein RB (1973) Determination of the frequency response of isometric soleus muscle in the cat using random nerve stimulation. J Physiol 229(2):275–296

Meek SG, Wood JE, Jacobsen SC (1990) Model-based, multimuscle EMG control of upper-extremity prostheses. In: Winters JM, Woo S-Y (eds) Multiple muscle systems. Springer, Berlin Heidelberg New York, pp 360–376

Mussa-Ivaldi FA, Hogan N, Bizzi E (1985) Neural, mechanical, and geometric factors subserving arm posture in humans. J Neurosci 5(10):2732–2743

Osu R, Franklin DW, Kato H, Gomi H, Domen K, Yoshioka T, Kawato M (2001) Short-and long-term changes in joint co-contraction associated with motor learning as revealed from surface EMG. J Neurophysiol 88(2):991–1004

Osu R, Gomi H (1999) Multijoint muscle regulation mechanisms examined by measured human arm stiffness and EMG signals. J Neurophysiol 81(4):1458–1468

Saltzman E (1986) Task dynamic coordination of the speech articulators: a preliminary model. Experimental Brain Research Series, vol 15. Springer, Berlin Heidelberg New York

Shaiman S (1989) Kinematic and electromyographic responses to perturbation of the jaw. J Acoust Soc Am 86(1):78–88

Stevens KN (1998) Acoustic phonetics. MIT Press, Cambridge, MA

Tremblay S, Shiller D, Ostry D (2003) Somatosensory basis of speech production. Nature 423:866–869