A model of facial biomechanics for speech production

Jorge C. Lucero[a)]
Departamento de Matematica, Universidade de Brasilia, Brasilia DF 70910-900, Brazil

Kevin G. Munhall[b)]
Departments of Psychology & Otolaryngology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 10 September 1998; revised 18 June 1999; accepted 23 June 1999)

Modeling the peripheral speech motor system can advance the understanding of speech motor control and audiovisual speech perception. A 3-D physical model of the human face is presented. The model represents the soft tissue biomechanics with a multilayer deformable mesh. The mesh is controlled by a set of modeled facial muscles which uses a standard Hill-type representation of muscle dynamics. In a test of the model, recorded intramuscular electromyography (EMG) was used to activate the modeled muscles and the kinematics of the mesh was compared with 3-D kinematics recorded with OPTOTRAK. Overall, there was a good match between the recorded data and the model's movements. Animations of the model are provided as MPEG movies. © *1999 Acoustical Society of America.* [S0001-4966(99)02810-6]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Jt [AL]

INTRODUCTION

The human face provides visible information during speech (Summerfield, 1992) and influences the acoustics of speech by determining the shape and size of the opening of the acoustic tube produced by the vocal tract (Lindblom and Sundberg, 1971). In recent years, there has been considerable interest in simulations of facial motion for the purposes of understanding speech motor control (e.g., Muller, Milenkovic, and McCleod, 1984), for producing realistic facial animation (Terzopoulos and Waters, 1990; Parke and Waters, 1996), and for stimulus generation for audiovisual speech research (Cohen and Massaro, 1990). In the present paper, we describe work on a 3-D facial model that extends the work of Terzopoulos and Waters (1990) on facial animation and produces a facial model that can be useful for speech perception and production research. In Terzopoulos and Waters' facial model, the biomechanical parameters related to muscles and skin, as well as geometrical dimensions, were selected using a heuristic approach. Although they were based on the actual physiology of a face, they were treated as dimensionless parameters, and their orders of magnitude were chosen so as to produce a realistic simulation. This approach complicates comparisons with experimental data. Here, we have tuned the model with realistic parameters obtained from experimental measurements. Further, we have modified the muscle geometry and the muscle model, according to physiological data. In addition, we have modified the manner in which motion is simulated using this model. In the original Terzopoulos and Waters' model, the face motion was obtained as a sequence of equilibrium states of the model. That is, at each single step (frame) of the animation, muscle forces were manually adjusted and the model was allowed to reach an equilibrium state before going to the next step. Although this technique may be used to produce reasonable animations, it does not correspond to the actual dynamics of the face. In the present work, we have driven the model dynamically, using perioral electromyographic data and 3-D position data recorded during speech production.

The present work follows the pioneering work of Eric Muller (Muller *et al.*, 1984) on facial modeling for physiological research. Muller argued that detailed modeling of the peripheral motor system is essential to understand the neural control of speech. A realistic representation of tissue and muscle permits control processes to be examined with the transfer function of the biological plant taken into account. There is abundant evidence that the peripheral motor system is not simply a passive channel for the transmission of signals from the central nervous system. Rather, the nonlinear mechanics of tissue and muscle, the inertial forces of the moving articulators, and the complexities of force generation in muscles perform a transform on those signals. The final form of the speech motor output is, thus, an interaction of the biomechanics and physiology of the vocal tract and the neural control signals.

The present research also has a second rationale. In audiovisual speech perception research, the visual stimuli are usually not controlled in any systematic fashion (see Munhall and Vatikiotis-Bateson, 1998). In published work in this area, it is rare to be provided with stimulus parameters for the moving face beyond the gender of the talker. This lack of direct visual stimulus control leaves many audiovisual experiments confounding image displacement and velocity factors with phonetic manipulations. Our secondary aim is to provide a tool that can be used to produce realistic facial animation in which facial movements can be manipulated in a systematic way for perception experiments (cf. Cohen and Massaro, 1990).

For both of these goals, the physics-based animation begun by Keith Waters and Demetri Terzopoulos (Lee, Terzopoulos, and Waters, 1993, 1995; Parke and Waters, 1996; Terzopoulos and Waters, 1993; Waters and Terzopoulos,

[a)] Electronic mail: lucero@mat.unb.br
[b)] Electronic mail: munhallk@psyc.queensu.ca

1991, 1992) offers a suitable research framework. The graphics environment created by Waters and Terzopoulos and their students allows physiological parameters for skin and muscle to be specified and permits realistic equations of motion to be implemented. This approach is consistent with a growing body of physiological modeling in speech which has proceeded on an articulator-by-articulator basis. Considerable progress has been made in modeling of the biomechanics of the vocal folds (Titze and Talkin, 1979), tongue (Kakita, Fujimura, and Honda, 1985; Wilhelms-Tricarico, 1995), velum (Berry, Moon, and Kuehn, 1998), jaw (Laboissière, Ostry, and Feldman, 1996) and tongue/jaw system (Sanguinetti, Laboissière, and Ostry, 1998). In each of these models, the biophysics of passive tissue as well as active muscle has been represented in great detail.

In order to model the human face in detailed biomechanical and physiological terms, a vast array of muscle properties and passive tissue characteristics must be specified. Unfortunately, good estimates are not available for all of these parameters. In spite of a long history of research interest in facial anatomy (Lightoller, 1925), there is still some uncertainty about the gross anatomy of the perioral musculature (e.g., Vinkka-Puhakka, Kean, and Heap, 1989) and little statistical data reporting the distribution of muscle lengths, muscle cross-sectional areas, etc. in the population. All of the facial muscles, and the perioral muscles in particular, are highly interdigitated (Blair, 1986; Blair and Smith, 1986), thus complicating their anatomical description. There is even less information about the motor unit/fiber types in the perioral muscles (cf. Sufit *et al.*, 1984).

The biomechanical properties of skin and facial tissues are also difficult to characterize. The constitutive equations for skin vary widely in the literature and parameters differ for different sites of the body, age, degree of obesity, etc. (See Lanir, 1987, for a review of skin modeling.) Further, the skin's properties vary according to direction. For example, the resting tension of the skin follows reliable directional patterns called Langer's lines (Barbenel, 1989).

I. FACIAL MODEL

This complex facial physiology is represented in our model by separate skin and muscle elements. The muscles are modeled using a standard Hill-type formulation (Winters, 1990; Zajac, 1989) that contains force generation due to the contractile element (the dependence of force on muscle length and velocity) and the passive dependence of force on muscle length. For a first approximation, we have assumed simple lines of action of the muscles and standard skeletal muscle physiology. With the exception of the orbicularis oris superior (OOS) and the orbicularis oris inferior (OOI), the perioral muscles have origins in the bony surfaces of the mandible and maxilla (see Kennedy and Abbs, 1979, for an overview of speech muscle anatomy). Thus, we have represented these muscles as linear force vectors. For the skin and connective tissues we have made similar first approximations. While the stress/strain characteristics of the skin are nonlinear and anisotropic (e.g., Lanir, 1987; Ho *et al.*, 1982; Larrabee, 1986), we have adopted a linear, isotropic approximation to the skin's mechanical characteristics. The skin is represented by a multilayered mesh that is parametrized with linear or piecewise linear estimates of the biomechanical properties of the skin. Finally, the facial morphology is individualized to match subjects using data from a laser range finder. This step allows direct comparisons between model behavior and recorded kinematics. Below, we provide the details for each component of the model.

A. Facial mesh

The modeled face consists of a deformable multilayered mesh. The nodes in the mesh are point masses, and each segment connecting nodes in the mesh consists of a spring and a damper in a parallel configuration. The nodes are arranged in three layers representing the structure of facial tissues. The top layer represents the epidermis, the middle layer represents the fascia, and the bottom layer represents the skull surface. The elements between the top and middle layers represent the dermal-fatty tissues, and elements between the middle and bottom layer represent the muscle. The skull nodes are fixed in the three-dimensional space. The fascia nodes are connected to the skull layer except in the region around the upper and lower lips and the cheeks.

The mesh has a uniform thickness with a separation of 1.5 mm between the topmost and middle layers and 2.5 mm between the middle and bottom layers.[1] All the nodes in the mesh have the same mass. Taking a mean skin density of 1142 kg/m^3 (Duck, 1990), and estimating from the model a mean node density of 5 node/cm^3, we obtain a mass $m = 0.23$ g for each node.

All springs, except for the dermal-fatty springs, are linear at elongation. We consider a Young's modulus for the skin of 7350 dyne/cm (Larrabee, 1986), and estimate the number of springs working in parallel in 1 cm^2 of mesh surface. Thus, we obtain a mean stiffness coefficient of about 600 dyne/cm for a spring 1 cm long. The stiffness coefficients of springs in the topmost layer are made higher (1200 dyne/cm) to represent the stiffer characteristic of the epidermis. The stiffness coefficients of all other springs are set to 600 dyne/cm. Since in general, the spring lengths are different than 1 cm, the stiffness coefficients for the actual springs in the mesh are properly scaled according to their rest length.

For the dermal-fatty springs, a biphasic approximation for the force-elongation characteristics is used (Parke and Waters, 1996; Terzopoulos and Waters, 1990). In real dermal tissue, the stiffness of the dermis with small stretches is mainly determined by elastin fibers, hence, the stiffness is low. As the elongation increases, collagen fibers uncoil. Once the collagen fibers are fully stretched, the skin stiffness increases suddenly and resists further elongation. The biphasic characteristic responds to the expression

$$g=\begin{cases} k_1 \Delta l, & \text{if } \Delta l/l_0 \leqslant 0.2 \\ 0.2k_1 l_0 + k_2(\Delta l - 0.2 l_0), & \text{if } \Delta l/l_0 > 0.2, \end{cases} \tag{1}$$

where g is the spring force, l_0 is the rest length, Δl is the elongation, and k_1 and k_2 are the stiffness coefficients. We adopt the estimated value of 600 dyne/cm for k_1 and 6000 dyne/cm for k_2 (for a spring with a rest length $l_0 = 1$ cm). The value of k_2 was set at 10 times the value of k_1 to ap-

TABLE I. Biomechanical constants of the facial mesh.

Parameter	Value
Mass	0.23 g
Damping	30 dyne s/cm
Stiffness:	
epidermal layer	1200 dyne/cm
dermal-fatty layer	600 dyne/cm (low deformation) 6000 dyne/cm (large deformation)
fascia layer	600 dyne/cm
muscle layer	600 dyne/cm

proximate the nonlinear function of the epidermal skin layer (Lanir, 1987).

At compression of the springs, we use the following nonlinear function to provide an infinite growth of the spring force as its length tends to zero (Lee *et al.*, 1995):

$$g = k \tan\left(\frac{\pi\sigma\Delta l}{2l_0}\right), \tag{2}$$

where k_1 is the same stiffness coefficient adopted for the elongation characteristics, and $\sigma = 0.98$ is a scaling factor.

The damping coefficient is $r = 30$ dyne s/cm for all the layers. This value was selected through visual evaluation of the animations. With a stiffness coefficient for the dermis $k = 600$ dyne/cm, the response time is $\tau = r/k = 50$ ms, which is in the order of experimental values (e.g., Muller *et al.*, 1984).

The above biomechanical constants for the skin are summarized in Table I.

B. Muscle models

The mesh is deformed by action of a set of modeled muscles of facial expression. The human face is controlled by dozens of anatomically distinct muscles, but a subset of 15 pairs of muscles is represented in the model. These modeled muscles can be divided into those muscles associated with upper face movement (corrugator, corrugator supercilli, major frontalis, lateral frontalis, inner frontalis) and the perioral muscles (depressor anguli oris, zygomatic major, zygomatic minor, levator labii superioris, levator labii nasi, depressor labii inferioris, risorius, mentalis, orbicularis oris superior, and orbicularis oris inferior). This subset of 15 muscles was chosen based on traditional analysis of emotional expression (Duchenne, 1990; Ekman and Friesen, 1975) and anatomical studies of the speech musculature (Kennedy and Abbs, 1979). Figure 1 shows the lines of action of these muscles.

All of the muscles, except the orbicularis oris superior and inferior, attach at one or more nodes of the fascia layer (middle layer). When activated, they exert a force on those nodes in the direction of the nodes of attachment to the skull layer [see Fig. 2(a)]. The orbicularis oris muscles attach to a path of fascia nodes along their length. When activated, they exert forces on the fascia nodes in the direction of that path [see Fig. 2(b)]. The nodes of attachment of the muscles were selected following anatomical descriptions in the literature (e.g., Kennedy and Abbs, 1979) and cadaver dissections carried out at Queen's University.

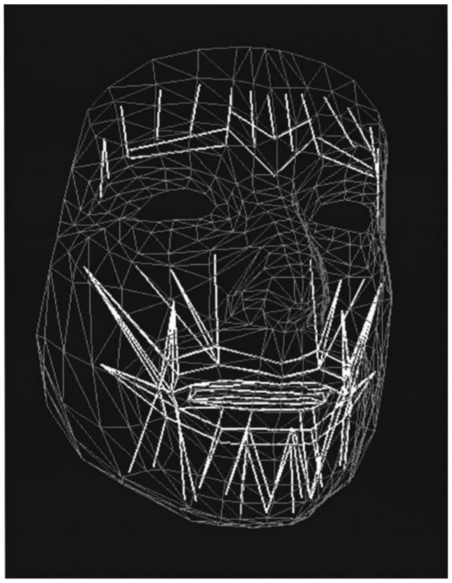

FIG. 1. Lines of action of facial muscles.

The generation of muscle force is computed by using integrated EMG as a measure of muscle activity, as follows.

The steady-state force $\overline{M}$ generated by the muscle is

$$\overline{M} = k_f S E, \tag{3}$$

where S is the muscle cross-sectional area, E is the integrated EMG level normalized to a range between 0 (mean of baseline muscle activity) and 1 (maximum activity recorded across the experiment, including a series of ''maximal'' facial gestures; cf. Zajac, 1989), and $k_f = 2500$ dyne/cm^2 is a

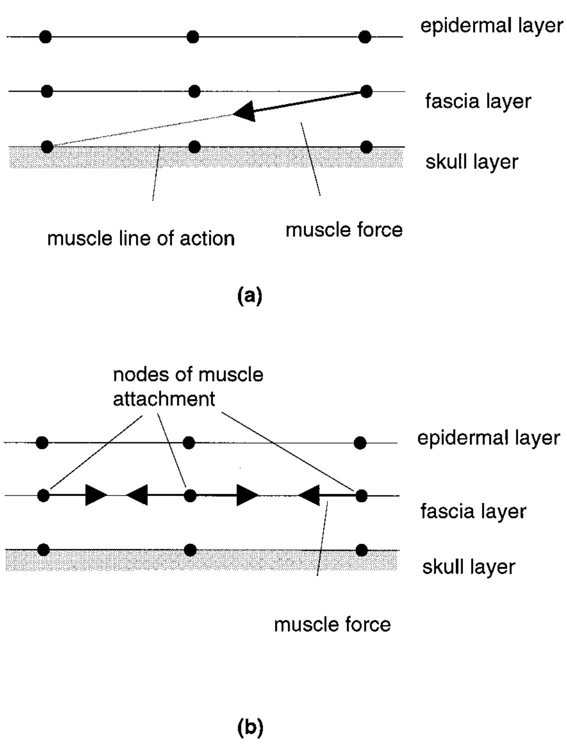

FIG. 2. Muscle force on the mesh. (a) Muscles attached to the skull, (b) orbicularis oris muscles.

TABLE II. Cross-sectional areas, stiffness, and number of fascia attachment for each muscle.

Muscle	Area (cm^2)	Stiffness (dyne/cm)
Zygomatic major	0.1	1730
Levator labii superioris	0.15	2595
Depressor anguli oris	0.4	6920
Depressor labii inferioris	0.11	1903
Mentalis	0.07	1211
Levator anguli oris	0.1	1730
Orbicularis oris superior	0.6	10 380
Orbicularis oris inferior	0.6	10 380

scaling coefficient (selected according to the results of the animations).

A graded force development of the muscle force M is simulated by the second-order, low-pass filtering of the steady-state force $\overline{M}$, according to the equation (Laboissière *et al.*, 1996)

$$\tau^2 \ddot{M} + 2\tau \dot{M} + M = \overline{M}, \tag{4}$$

where $\tau = 15$ ms. A force-length characteristic is added using the equation (Otten, 1987; Brown, Scott, and Loeb, 1996)

$$M' = M \exp\left[- \left| \frac{(l/l_0)^{2.3} - 1}{1.26} \right|^{1.62} \right], \tag{5}$$

where l is the actual muscle length and l_0 its rest length.

Finally, force-velocity and passive stiffness characteristics are added to compute the total muscle force F, according to the equation (Laboissière *et al.*, 1996)

$$F = M[f_1 + f_2 \arctan(f_3 + f_4 \dot{l})] + [k_m \Delta l]^+, \tag{6}$$

where $f_1 = 0.82$, $f_2 = 0.5$, $f_3 = 0.43$, $f_4 = 0.2$ s/cm, k_m is the passive muscle stiffness, and

$$[x]^+ = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leqslant 0. \end{cases} \tag{7}$$

The passive muscle stiffness for each muscle was computed by scaling a reference value according to the cross-sectional area of each muscle. As reference, we used a cross-sectional area of 1 cm^2 and passive stiffness of 17 300 dyne/cm for the hyoid depressor muscle (Laboissière *et al.*, 1996). The cross-sectional areas were taken from Kennedy and Abbs (1979) or estimated from experimental measurements on a dissected cadaver. Table II shows the cross-sectional areas and the passive stiffness used in the simulations. Only the perioral muscles shown in the table were considered for the present work.

The steady-state muscle force was also computed by scaling according to the cross-sectional area, as explained later in Sec. II C.

C. Equations of motion

The equation of motion for each node i of the model has the general expression (Lee *et al.*, 1995)

$$m \frac{d^2 \mathbf{x}_i}{dt^2} + r \sum_j \left(\frac{d\mathbf{x}_i}{dt} - \frac{d\mathbf{x}_j}{dt} \right) + \sum_j \mathbf{g}_{ij} + \sum_e \mathbf{q}_i^e + \mathbf{s}_i + \mathbf{h}_i = \mathbf{F}_i . \tag{8}$$

In this equation, $\mathbf{x}_i$ is the current position of node i. The second term is the total damping force acting on the node, and the index j represents all the nodes that are neighbor to node i. The third term is the total spring force, and the force contribution $\mathbf{g}_{ij}$ of spring-connecting nodes i and j is calculated using Eqs. (1) and (2).

The fourth term models the incompressibility of human skin. $\mathbf{q}_i^e$ is the force at node i, produced by the preservation of volume of the triangular prism element e to which node i belongs. This force is calculated as

$$\mathbf{q}_i^e = k_{e1}(V^e - \tilde{V}^e)\mathbf{n}_i^e + k_{e2}(\mathbf{p}_i^e - \tilde{\mathbf{p}}_i^e), \tag{9}$$

where V^e and $\tilde{V}^e$ are, respectively, the current and rest volumes of element e, $\mathbf{n}_i^e$ is the epidermal normal at node i, $\mathbf{p}_i^e$ and $\tilde{p}_i^e$ are the current and rest nodal coordinates for node i with respect to the center of mass of element e, and $k_{e1} = 1000$ dyne/cm^3, $k_{e2} = 2000$ dyne/cm are scaling factors.

The fifth term $\mathbf{s}_i$ in Eq. (8) is a force to penalize fascia nodes penetrating the skull. This force cancels out the force component on the fascia node in the direction towards the skull, and is calculated as

$$\mathbf{s}_i = \begin{cases} -(\mathbf{f}_i^n \cdot \mathbf{n}_i)\mathbf{n}_i, & \text{if } \mathbf{f}_i^n \cdot \mathbf{n}_i < 0 \\ 0, & \text{otherwise} \end{cases}, \tag{10}$$

where $\mathbf{f}_i^n$ is the net force on fascia node i and n_i is the nodal normal.

The last term on the left side in Eq. (8), $\mathbf{h}_i$, is a nodal restoration force applied to the fascia nodes connected to the skull. It is calculated as

$$\mathbf{h}_i = k_h(\mathbf{x}_i - \tilde{\mathbf{x}}_i), \tag{11}$$

where $\tilde{\mathbf{x}}_i$ is the rest position of fascia node i and $k_h = 200$ dyne/cm is a scaling factor. This equation acts as an extra force modeling the attachment to the skull of the skin, and compensates in part the cancellation of the force component between fascia nodes and the skull due to penalization of skull penetration. It is necessary to help bring the nodes back to the rest (initial) position when muscle forces are deactivated (without this force, the nodes tend to wander around the rest position).

Finally, $\mathbf{F}^i$ in Eq. (8) is the total muscle force applied to node i.

II. FACIAL ANIMATIONS

The model described above represents a first approximation of the peripheral biomechanics and physiology of the human face. To test the accuracy of this representation of the plant, electromyographic (EMG) data were collected from a set of seven perioral muscles. The aim was to test the transfer function between muscle activity and facial surface kinematics and to examine the model's capability to reproduce the dynamical behavior of the face during speech production. Specifically, we used the recorded EMG to drive the modeled muscles. We then compared the model kinematics to the observed subject kinematics. The model was individualized to the subject's morphology using data from a Cyberware laser scanner (Lee *et al.*, 1993, 1995). Thus, direct kinematic

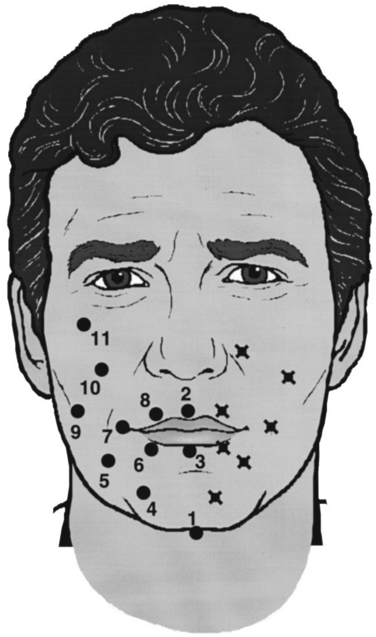

FIG. 3. Position of OPTOTRAK IREDs and electrode insertion points (crosses) for EMG collection.

comparisons were possible. The next sections will describe in detail the animation process and comparisons with recorded facial kinematics.

A. EMG and kinematic data

Intramuscular EMG data were collected from perioral muscles on the left side of a single subject's face, while the subject produced English sentence material [20 Central Institute of the Deaf (CID) everyday sentences].[2] The subject was a male, native speaker of American English. The EMG was recorded from the levator labii superioris, levator anguli oris/zygomatic major,[3] depressor anguli oris, depressor labii inferioris, mentalis, orbicularis oris superior, and orbicularis oris inferior using intramuscular hooked-wire, bipolar electrodes. Figure 3 shows the approximate electrode positions (crosses). Electrode insertions were determined with reference to Kennedy and Abbs (1979) and were verified using a series of nonspeech maneuvers. The acoustic signal was simultaneously recorded. The sampling frequency of the EMG and acoustic signal was 2500 Hz.

At the same time, we recorded the three-dimensional position of 11 markers (infrared emitting diodes (IRED)) on the right side of the face (see Fig. 3) using an OPTOTRAK (model 3010, Northern Digital, Inc.) system at a sampling frequency of 60 Hz. The position data were corrected for motion of the head, and transformed to a coordinate system in which the origin is the incisor cusp and the horizontal and protrusion axes lie along the bite surface (Ramsay *et al.*, 1996).

B. Data preprocessing

The EMG signals were first rectified and next integrated and downsampled to 60 Hz to match the sampling frequency of the position data, using a median filtering algorithm with a 17-ms trapezoidal window (Vatikiotis-Bateson and Yehia, 1996). Finally, the signals for each muscle were normalized to a range between 0 and 1 by dividing them by the maximum level of each muscle. As indicated above, each muscle's maximum was set to the highest level recorded in the speech material or during a set of extreme facial gestures (e.g., extreme lip protrusion).

The position data from the OPTOTRAK system were transformed to the coordinate system used in the face model. In the face model, the origin is at the node immediately below the highest node of the nose, and the *x*-axis is horizontal from left to right, the *y* is vertical to the top, and the *z*-axis is the protrusion axis.

Since the dynamics of the jaw were not yet implemented in the model, we rotated the jaw using position data of the subject's chin during the animations for the CID sentences. The rotation of jaw was computed using the OPTOTRAK data for marker 1. First, the vertical displacement of the marker was computed, in relation to its rest (initial) position. Then, we computed the rotation angle of the nodes in the jaw in the facial mesh that would produce the same vertical displacement of these nodes. In the case of the bite-block experiments, the jaw rotation was kept fixed at its initial value, computed from the OPTOTRAK data.

C. Animation

The face model was implemented as a set of programs written in C language and using OPENGL for the graphic interface, adapted from the original programs by Lee *et al.* (1995). It runs on an Ultra Sparc workstation, and an animation of 3 s took about 4 min to compute.

The animation was performed as follows. The equations of motion of the mesh nodes were solved with an Euler algorithm, and a time step of 0.33 ms. To obtain a final rate of 60 Hz, a frame with the animated face was saved every 50 iterations of the algorithm. Also, the positions of nodes closest to the positions of markers in the subject's face were saved every 50 iterations.

At the beginning of each series of 50 iterations, the computed rotation of the jaw was read, and all the jaw nodes in the mesh at the skull layer were rotated accordingly. To compensate for a time delay in the propagation of the jaw rotation from the skull layer to the epidermal layer (recall that jaw rotation was computed from a marker on the epidermis), we introduce an artificial time advance of two sampling points (33 ms) to the jaw rotation data. Next, the processed EMG was read, and the force exerted by each muscle was computed. The activity level of the zygomatic major was set equal to the levator anguli oris.

The equations of motion were then solved, considering the muscle force and the jaw rotation constants during the 50 iteration period. This process was repeated until the end of the EMG data files.

D. Results for CID sentences

Figures 4, 5, and 6 show the displacement (vertical and protrusion)[4] and acceleration of nodes corresponding to IREDs 3, 5, and 7, for the CID sentence ''Where are you going?'' The figures also show the measured displacement

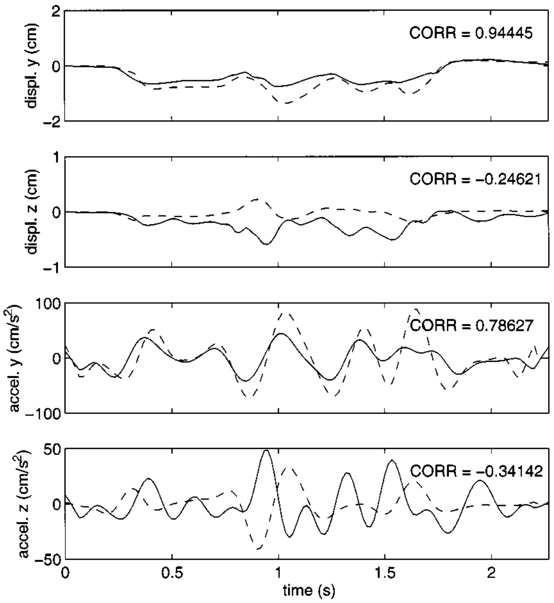

FIG. 4. Displacement and accelerations in the vertical (y) and protrusion (z) directions corresponding to IRED 3 for the CID sentence ''Where are you going?'' Full line: animation results; broken line: measured data. The cross-correlation between the animation and measured data is shown.

and acceleration of the IREDs, and the cross-correlation between the animated and measured kinematics.

In general, there is a good match between the animated and measured kinematics. Tables III and IV show mean, maximum, and minimum cross-correlation coefficients for all IREDs and corresponding model nodes for all of the sentences. As can be seen, the match tends to be better in the vertical (y) displacements and acceleration than the protrusion (z) records. There is also a difference in the degree of correlation across the various IRED positions. There is a tendency for the IREDs immediately surrounding the mouth to show lower cross-correlations.[5]

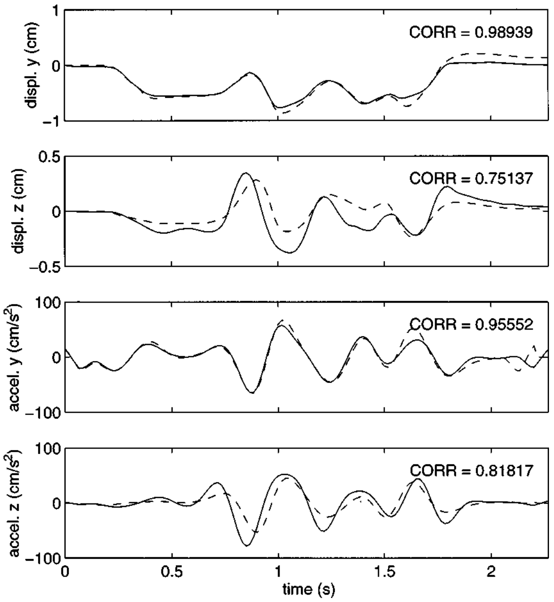

FIG. 5. Displacement and accelerations in the vertical (y) and protrusion (z) directions corresponding to IRED 5 for the CID sentence ''Where are you going?'' Full line: animation results; broken line: measured data. The cross-correlation between the animation and measured data is shown.

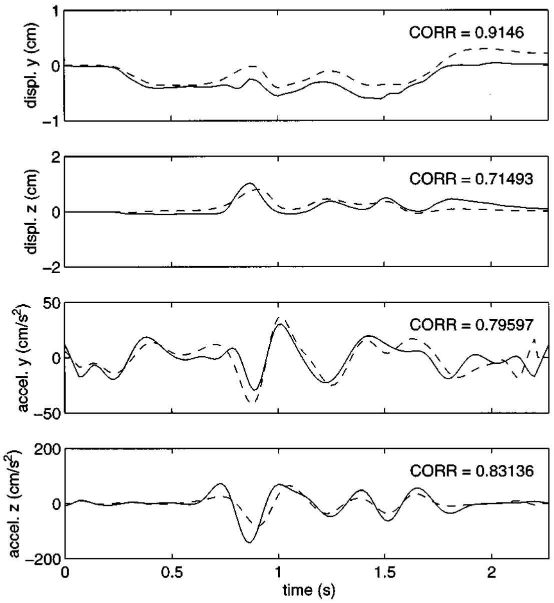

FIG. 6. Displacement and accelerations in the vertical (y) and protrusion (z) directions corresponding to IRED 7 for the CID sentence ''Where are you going?'' Full line: animation results; broken line: measured data. The cross-correlation between the animation and measured data is shown.

E. The movies

Research on the face allows a second type of measure of the success of modeling efforts. In addition to statistical measures of movement similarity between the model and real facial kinematics, the animations can be simply viewed to assess the degree of perceived realism of the motion. Two MPEG movies are available on our web page[6] for this purpose. The first shows the CID sentence ''Where are you going?'' The second shows repetition of the vowel–consonant–vowel (VCV) utterance /upæ/ with the subject using a bite block to immobilize the jaw. EMG was collected for the same muscle set and used to drive the model. Thus, for the bite-block movie all of the animation is produced by the muscle activation. The movies are produced at 60 frames per s and their viewing speed will depend on the processor

TABLE III. Cross-correlation coefficients between animated and measured displacements of markers for CID sentences.

	Vertical			Protrusion		
Marker #	Mean	Maximum	Minimum	Mean	Maximum	Minimum
1	0.986	0.996	0.969	0.840	0.932	0.578
2	0.344	0.645	0.072	−0.001	0.581	−0.617
3	0.896	0.954	0.775	0.369	0.772	−0.246
4	0.976	0.992	0.949	0.739	0.883	0.307
5	0.971	0.991	0.941	0.694	0.917	0.523
6	0.916	0.963	0.831	0.265	0.600	−0.428
7	0.886	0.958	0.728	0.469	0.770	0.087
8	0.508	0.883	−0.315	0.309	0.816	0.068
9	0.857	0.930	0.747	0.724	0.930	0.566
10	0.467	0.714	−0.063	0.371	0.785	−0.199
11	0.785	0.951	0.515	0.294	0.561	−0.104

TABLE IV. Cross-correlation coefficients between animated and measured accelerations of markers for CID sentences.

	Vertical			Protrusion		
Marker #	Mean	Maximum	Minimum	Mean	Maximum	Minimum
1	0.959	0.985	0.851	0.870	0.945	0.657
2	0.339	0.600	−0.086	−0.121	0.420	−0.674
3	0.848	0.933	0.601	0.263	0.625	−0.341
4	0.942	0.970	0.844	0.714	0.881	0.184
5	0.946	0.976	0.835	0.702	0.868	0.307
6	0.861	0.942	0.670	0.331	0.778	−0.071
7	0.867	0.957	0.662	0.505	0.831	0.112
8	0.408	0.765	−0.490	0.288	0.787	−0.267
9	0.773	0.926	0.375	0.661	0.873	0.336
10	0.365	0.768	−0.314	0.291	0.735	−0.300
11	0.737	0.851	0.539	0.410	0.690	−0.108

speed of the reader's computer. As can be seen, in both cases the model produces highly realistic speech movement and natural skin motion.

III. DISCUSSION

The model described here incorporates active muscle properties as well as passive muscle and tissue properties in a detailed 3-D simulation of facial dynamics. Realistic speech animation is produced by driving the physical model with recorded EMG data. Good cross-correlations between model kinematics and recorded data were observed and natural patterns of skin deformation can be observed in movies of the animation. This initial version of the model is quite promising, yet both the model and the test of the model involved significant simplifications. The tissue biomechanical properties are represented by linear approximations. While the muscle activation dynamics are represented in a sophisticated manner, the lines of action of the muscles are simplified. In addition, only a subset of the full set of facial muscles is modeled.

The test of the model's performance was shaped by a number of practical considerations. The EMG and kinematics were recorded from opposite sides of the face. This was done to avoid having electrical noise from the OPTOTRAK contaminate the muscle activity recordings. Thus, we have tacitly assumed symmetry in the structure and action of the face. This assumption ignores the known asymmetries in facial morphology and lip movement (Campbell, 1982) and thus adds error variance to our modeling. Our use of the EMG to drive the model allowed a direct test of the representation of the biomechanics of the face. However, the use of perioral EMG raises a set of separate issues. Intramuscular EMG recordings such as the ones we used are imperfect measures of the full muscle activation and force generation. The reasons for this include recording noise in the signals, interdigitation of the muscle fibers potentially leading to recordings from multiple muscles at any single recording site (Blair and Smith, 1986), possible compartmentalized muscles (Binder and Stuart, 1980) in which different motor units within a muscle have different functional roles, and nonlinearities between EMG and force generation.

In spite of these potential problems, the model's performance was surprisingly good. What accounts for the model performance? No single factor can explain this, but a combination of the following factors seems most likely. The primary determinant of facial motion is the movement of the jaw. When the jaw opens, the facial tissue and muscles are stretched and the skin deforms to accommodate the movement. In the simulations, we moved the jaw based on recorded kinematics and thus the modeled facial tissue responded well to this change. While the tissue changes in response to the jaw kinematics are realistic, it leaves the question of the extent to which the perioral muscles are accurately portrayed. The bite-block trial shown in the second movie indicates that realistic animation can be produced in the absence of jaw movement. However, we did not collect enough of the bite-block data to carry out statistical analyses. It is likely the cross-correlations would be lower in this case.

A second contribution to the good performance is the fact that the face in speech is controlled with few degrees of freedom (Ramsay *et al.*, 1996) and may be quite crudely controlled (Löfqvist and Gracco, 1997). Ramsay *et al.*'s principal component analysis of the lip motion indicates that the motion along a single dominant trajectory accounted for much of the variance in the data and that the motion of any single position marker on the lip was strongly one-dimensional. Löfqvist and Gracco have shown that the lips often make contact in bilabial stops at peak velocity and that contact forces are involved in deceleration. This would imply that there is considerable redundancy in the motor control of the lips and face for phonetic targets and that these targets are not specified with great precision. Thus, with tissue parameters in the biological ballpark and EMG patterns prescribing a time structure for the behavior, the facial animation looked realistic.

As noted above, the potential for problems in the EMG is great and its use is a bit of an experimental gamble. However, the performance of the model here and separate analyses on the same EMG and kinematic data (Vatikiotis-Bateson and Yehia, 1996) indicate that the recorded EMG signals were very good correlates of the muscle activity. Vatikiotis-Bateson and Yehia (1996) have shown, using a second-order autoregressive model, that the EMG can be used to estimate the facial motion with very high accuracy. One interpretation of this is that the facial muscle activity overdetermines the relatively simple facial speech gestures. Thus, in combination, a range of muscle recordings can provide good estimates of the time course of perioral force generation.

One final aspect of the model may contribute to its good performance. The model's overall performance may be dominated by the mesh viscoelastic properties. If this is the case, the response of the model will be determined mainly by the time constants of the mesh, and high accuracy in the time histories of the muscle activities would not be necessary. We are currently exploring this question with sensitivity analysis of the model.

Models of this kind provide an essential tool for understanding speech motor control. In the study of speech, we can measure only the end product of a complex chain of planning and control processes. However, the kinematics of speech and the EMG are the result of an interaction between linguistic and motoric planning processes and the biomechanics of the speech articulators. To understand this complex sensorimotor process, we must be able to assign variance components in the data to different stages in the production process. At the very least, variance components due to the central commands and the biological plant must be separated. If models such as the one described here can provide a biologically plausible representation of the plant, then comparisons between different ideas about the central control of the speech motor system can be made.

This particular model also has another role in speech research. As indicated in the introduction, there is little visual stimulus control or visual stimulus specification in research on audiovisual speech perception. Some individuals are easier to lipread than others, and speaking style and phonetic context change the facial kinematics. The timing and velocities of speech movements, the magnitudes of facial motions, the visibility of the oral cavity, and the size and velocity of head motion all can vary from talker to talker and from context to context. Yet, visual stimulus characteristics are rarely reported (cf. Munhall *et al.*, 1996; Munhall and Tohkura, 1998). In acoustic speech perception, the field has developed on the basis of detailed synthesis and multivariate parameter specification. Audiovisual and visual speech perception research must follow similar standards of stimulus control, and models such as the one described in this paper will be important tools for creating factorial studies of visual cues (cf. Massaro, 1987, 1998).

ACKNOWLEDGMENTS

We are grateful to Eric Vatikiotis-Bateson and Vince Gracco for help collecting the EMG and OPTOTRAK data. The data were collected at ATR Human Information Processing Research Laboratories and the authors gratefully acknowledge their support. We thank Demetri Terzopoulos and Victor Lee for generously sharing their time and software. André Tan helped with the cadaver dissection. Michel Pitermann made helpful comments on an earlier draft. This work was funded by NIH Grant No. DC-00594 from the National Institute of Deafness and other Communication Disorders and NSERC.

[1] This is a simplification that we will explore in future research. The skin's layers are not uniform in thickness (e.g., Kennedy and Abbs, 1979) and the location of tissue thickness changes presumably contributes significantly to individual facial characteristics.

[2] Data for two of the sentences were lost due to recording errors.

[3] We could not reliably distinguish these muscles and thus we have driven both of these muscles in the model with the signal from this single recording site.

[4] Since there is little lateral motion in speech (Vatikiotis-Bateson and Ostry, 1995), we have focused on the vertical and protrusion movements only. However, note that this is a 3-D model of the face with motion in all planes.

[5] The poorer performance of the model at the lips is most likely due to an omission in the current model. The lips do not penetrate each other when they make contact because of a penetration penalty force, but we have not modeled the friction forces on the lip surfaces. As a result, the lips tend to slide upon contact rather than compress and deform. This results in inaccuracies in the detailed kinematics of lip shape. In our current work, we are exploring implementation of a skin surface friction.

[6] http://130.15.96.12/~munhallk/home.html

Barbenel, J. C. (**1989**). ''Biomechanics of Skin,'' in *Systems and Control Encyclopedia: Theory, Technology, Applications* (Pergamon, Oxford).

Berry, D. A., Moon, J. B., and Kuehn, D. P. (**1998**). A Histologically-based Finite Element Model of the Soft Palate, *National Center for Voice and Speech: Status and Progress Report* (Vol. 12, pp. 71–77).

Binder, M. D., and Stuart, D. G. (**1980**). ''Motor Unit-muscle Receptor Interactions: Design Features of the Neuromuscular Control System,'' in *Progress in Clinical Neurophysiology, Spinal and Supraspinal Mechanisms of Voluntary Motor Control and Locomotion*, edited by J. E. Desmedt (Karger, Basel, Switzerland).

Blair, C. (**1986**). ''Interdigitating Muscle Fibers Throughout Orbicularis Oris Inferior: Preliminary Observations,'' J. Speech Hear. Res. **29**, 266–269.

Blair, C., and Smith, A. (**1986**). ''EMG Recording in Human Lip Muscles: Can Single Muscles Be Isolated?'' J. Speech Hear. Res. **29**, 256–266.

Brown, I. E., Scott, S. H., and Loeb, G. E. (**1996**). ''Mechanics of Feline Soleus: II. Design and Validation of a Mathematical Model,'' J. Muscle Res. Cell Motil. **17**, 221–233.

Campbell, R. (**1982**). ''Asymmetries in Moving Faces,'' British J. Psychol. **73**, 95–103.

Cohen, M. M., and Massaro, D. W. (**1990**). ''Synthesis of Visible Speech,'' Behav. Res. Methods Instrum. Comput. **22**, 260–263.

Duchenne, G. B. (**1990**). *The Mechanism of Human Facial Expression* (Cambridge University Press, New York).

Duck, F. A. (**1990**). *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, London).

Ekman, P., and Friesen, W. V. (**1975**). *Unmasking the Face* (Consulting Psychologists, Palo Alto, CA).

Ho, S. P., Azar, K., Weinstein, S., and Bowley, W. W. (**1982**). ''Physical Properties of Human Lips: Experimental and Theoretical Analysis,'' J. Biomech. **15**(11), 859–866.

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). ''Computation of Mapping from Muscular Contraction Patterns to Format Patterns in Vowel Space,'' in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, New York).

Kennedy, J. G., and Abbs, J. H. (**1979**). ''Anatomic Studies of the Perioral Motor System: Foundations for Studies in Speech Physiology,'' Speech Lang. Adv. Res. Pract. **1**, 211–270.

Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). ''The Control of Multimuscle Systems: Human Jaw and Hyoid Movements,'' Biol. Cybern. **74**, 373–384.

Lanir, Y. (**1987**). ''Skin Mechanics,'' in *Handbook of Bioengineering*, edited by R. Skalak and S. Chien (McGraw-Hill, New York).

Larabee, W. F. (**1986**). ''A Finite Element Model of Skin Deformation. I. Biomechanics of Skin and Soft Tissue: A Review,'' Laryngoscope **96**, 399–405.

Lee, Y., Terzopoulos, D., and Waters, K. (**1993**). Constructing Physics-based Facial Models of Individuals. Paper presented at the Proceedings of Graphics Interface '93.

Lee, Y., Terzopoulos, D., and Waters, K. (**1995**). ''Realistic Modeling for Facial Animation,'' Comput. Graph. **29**, 55–62.

Lightoller, G. H. S. (**1925**). ''Facial Muscles: The Modiolus and Muscles Surrounding the Rima Oris with Some Remarks About the Panniculus Adiposus,'' J. Anat. **LX** (Part 1), 1–85.

Lindblom, B., and Sundberg, J. (**1971**). ''Acoustical Consequences of Lip, Tongue, Jaw and Larynx Movement,'' J. Acoust. Soc. Am. **50**, 1166–1179.

Löfqvist, A., and Gracco, V. L. (**1997**). ''Bilabial Stop Consonant Production: Lip and Jaw Kinematics.'' J. Speech Hear. Res. **40**, 877–893.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye* (Erlbaum, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces* (Bradford, Cambridge, MA).

Muller, E. M., Milenkovic, P., and MacLeod, G. (**1984**). ''Perioral Tissue Mechanics During Speech Production,'' in *Proceedings of the Second IMAC International Symposium on Biomedical Systems Modeling*, edited by C. DeLisi and J. Eisendfeld (North Holland, Amsterdam).

Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (**1996**). ''Temporal Constraints on the McGurk Effect,'' Percept. Psychophys. **58**, 351–362.

Munhall, K. G., and Tohkura, Y. (**1998**). ''Audiovisual Gating and the Time Course of Speech Perception,'' J. Acoust. Soc. Am. **104**, 530–539.

Munhall, K. G., and Vatikiotis-Bateson, E. (**1998**). ''The Moving Face During Speech Communication,'' in *Hearing By Eye, Part 2: The Psychology of Speechreading and Audiovisual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Taylor & Francis Psychology, London).

Otten, E. (**1987**). ''A Myocybernetic Model of the Jaw System of the Rat,'' J. Neurosci. Methods **21**, 287–302.

Parke, F., and Waters, K. (**1996**). *Computer Facial Animation* (AK Peters, Wellesley, MA).

Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (**1996**). ''Functional Data Analyses of Lip Motion,'' J. Acoust. Soc. Am. **99**, 3718–3728.

Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''An Integrated Model of the Biomechanics and Neural Control of the Tongue, Jaw, Hyoid and Larynx Systems,'' J. Acoust. Soc. Am. **103**, 1615–1627.

Sufit, R. L., Poulsen, G., Welt, C., and Abbs, J. H. (**1984**). ''Morphology and Histochemistry of the Facial Muscles and Fascicularis,'' Soc. Neurosci. Abs. **10**, 779.

Summerfield, Q. (**1992**). ''Lipreading and Audio-Visual Speech Perception,'' Philos. Trans. R. Soc. London, Ser. B **335**, 71–78.

Terzopoulos, D., and Waters, K. (**1990**). ''Physically-based Facial Modeling, Analysis, and Animation,'' Visual. Comput. Anim. **1**, 73–80.

Terzopoulos, D., and Waters, K. (**1993**). ''Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models,'' IEEE Trans. Pattern. Anal. Mach. Intell. **15**, 569–579.

Titze, I. R., and Talkin, D. T. (**1979**). ''A Theoretical Study of the Effects of Various Laryngeal Configurations on the Acoustics of Phonation,'' J. Acoust. Soc. Am. **66**, 60–74.

Vatikiotis-Bateson, E., and Ostry, D. (**1995**). ''An Analysis of the Dimensionality of Jaw Motion in Speech,'' J. Phonetics **23**, 101–117.

Vatikiotis-Bateson, E., and Yehia, H. (**1996**). ''Physiological Modeling of Facial Motion During Speech,'' Trans. Tech. Com. Psycho. Physio. Acoust., H-96 **65**, 1–8.

Vinkka-Puhakka, H., Kean, M. R., and Heap, S. W. (**1989**). ''Ultrasonic Investigation of the Circumoral Musculature,'' J. Anat. **166**, 121–133.

Waters, K., and Terzopoulos, D. (**1991**). ''Modeling and Animating Faces Using Scanned Data,'' J. Visualiz. Comput. Anim. **2**, 123–128.

Waters, K., and Terzopoulos, D. (**1992**). ''The Computer Synthesis of Expressive Faces,'' Philos. Trans. R. Soc. London, Ser. B **335**, 87–93.

Wilhelms-Tricarico, R. (**1995**). ''Physiological Modeling of Speech Production: Methods for Modeling of Soft-Tissue Articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.

Winters, J. M. (**1990**). ''Hill-based Muscle Models: A Systems Engineering Perspective,'' in *Multiple Muscle Systems: Biomechanics and Movement Organization*, edited by J. Winters and S. Woo (Springer, London), pp. 69–93.

Zajac, F. E. (**1989**). ''Muscle and Tendon: Properties, Models, Scaling and Application to Biomechanics and Motor Control,'' Crit. Rev. Biomed. Eng. **17**, 359–411.