ELSEVIER

Speech Communication 22 (1997) 185–205

SPEECH COMMUNICATION

Synthesis of V–V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis [1]

Yohan Payan *, Pascal Perrier [2]

Institut de la Communication Parlée, UPRESA CNRS 5009, INPG & Université Stendhal Grenoble, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 01, France

Received 25 October 1996; revised 11 April 1997; accepted 28 April 1997

Abstract

An assessment of a target-based control model of speech production using Feldman's Equilibrium Point Hypothesis is presented. It consists of simulations of articulatory movements during Vowel-to-Vowel sequences with a 2D biomechanical tongue model. In the model the main muscles responsible for tongue movements and tongue shaping in the mid-sagittal plane are represented. The elastic properties are accounted through a Finite-Element modeling, while force generation principles are implemented according to the non-linear force-length Invariant Characteristics proposed by Feldman. Movement is produced through control variable shifts at rates that are constant throughout each transition. The external contours of the model are adjusted to approximate X-ray data collected on a native speaker of French, and it is inserted in the vocal tract contours of the speaker. Thus, from tongue shapes generated with the model, it was possible to produce formant trajectories compatible with the speaker's acoustic space. It permitted a comparison of simulations with real data collected on the speaker in the kinematic and acoustic domains. Emphasis is put on the realism of synthesized formant trajectories, and on the potential influence of biomechanical tongue properties on to measurable kinematic features. © 1997 Elsevier Science B.V.

Zusammenfassung

Wir stellen eine Bewertung eines Sprachproduktionsmodells vor, das das Konzept von phonetischem Ziel einschliesst, und auf der Equilibrium Point Hypothesis von Feldman beruht. Dafür wurden artikulatorische Bewegungen in vokalischen Reihenfolgen mit einem zweidimensionalen Zungenmodell simuliert. In diesem Modell wurden die Muskeln eingeschlossen, deren Wirkungen auf die Zungengestalt in der sagitalen Ebene die wichtigsten sind. Durch die Methode der endlichen Elemente wurden die elastischen Eigenschaften der Zunge modellisiert, und die muskuläre Kraft entspricht Feldmans unlinearen "Kraft-Länge unveränderlichen Kurven". Die Bewegung wurde durch die Veränderung der Kontrollparameter mit einer in jeder vokalischen Reihenfolge gleichbleibenden Geschwindigkeit produziert. Die Außenränder des Zungenmodells wurden Röentgenstrahldaten angepasst, die bei einem französischen Sprecher erworben wurden; das Modell wurde innerhalb der Vokaltraktränder dieses Sprechers integriert. Dadurch wurde es möglich Formantbahnen aus den zeitlichen Reihenfolgen von Zungengestalten zu synthetisieren, die den akustischen Eigenschaften des Sprechers entsprechen.

* Corresponding author. E-mail: payan@icp.grenet.fr.

[1] Videofiles available. See http://www.elsevier.nl/locate/specom .

[2] E-mail: perrier@icp.grenet.fr.

PII S0167-6393(97)00019-8

Simulationen und Daten wurden danach vergleichbar. Dieser Vergleich betont den Realismus der synthetisierten Formantbahnen, und zeigt inwiefern die biomekanischen Eigenschaften einer Sprache potentiell die auswertbaren kinematischen Charakteristika beeinflussen können. © 1997 Elsevier Science B.V.

Résumé

Nous présentons une évaluation d'un modèle de contrôle de la production de la parole exploitant la notion de cible, fondé sur l'Hypothèse du Point d'Équilibre de Feldman. Elle consiste en des simulations, avec un modèle biomécanique bi-dimensionnel de la langue, de mouvements articulatoires lors de transitions vocaliques. Dans le modèle, les principaux muscles agissant sur la forme de la langue dans le plan sagittal sont représentés. La méthode des éléments finis modélise les propriétes élastiques de l'articulateur, et les principes de génération de force sont conformes aux Caractéristiques Invariantes Force-Longueur non linéaires de Feldman. Le mouvement est produit en déplaçant les variables de contrôle à vitesse constante pour chacune des transitions. Les contours externes de la langue sont ajustés pour correspondre aux données radiographiques acquises sur un locuteur français, et le modèle est placé à l'intérieur des contours fixes du conduit vocal de ce locuteur. Ainsi, il a été possible de synthétiser, à partir des contours du modèle, des trajectoires formantiques adaptées à l'espace acoustique du locuteur, afin de comparer simulations et données. Nous insistons sur le caractère réaliste des trajectoires formantiques synthétisées, et sur le rôle potentiel des propriétés biomécaniques de la langue sur les caractéristiques cinématiques mesurables. © 1997 Elsevier Science B.V.

Keywords: Biomechanical modeling; Speech production; Speech motor control; Velocity profiles

1. Introduction

Interesting insights into speech control have been provided from studies based on analyses of articulatory and/or acoustic signals (Munhall et al., 1985; Nittrouer et al., 1988; Sock and Löfqvist, 1995). However, from our point of view, they suffer from an important flaw. The physical speech signals are, in fact, the consequence of the interaction between the mechanical structure of the articulatory apparatus and the forces that act on it. This interaction will be referred to as *dynamics* in the paper. Furthermore, part of these forces are the result of a control, while the other part depends on peripheral feedback. Thus, to understand the control of speech production from the articulatory or acoustic signals, it seems to us that it is necessary to model separately the *biomechanics*, i.e. the morphological and dynamical properties of the structure, and the neurophysiology of the control. Using a dynamical model of speech articulators, and assuming that speech control is target-based, we presented elsewhere how part of the measured speech variability could be explained by the physical properties of the system (Perrier et al., 1996a; Lœvenbruck, 1996). For that, the control of the model was based on Feldman's (1966) Equilibrium Point Hypothesis (EPH), which proposes that limb movements are produced by centrally specified shifts of the mechanical equilibrium of the peripheral motor system (see also similar results obtained with a biomechanical jaw model in Perrier et al., 1996b).

This paper presents simulations obtained with an original 2D biomechanical model of the tongue, whose control is based on the same principles. Articulatory and acoustic signals are compared to data collected on a native speaker of French. In the analysis, emphasis will be put on the contribution of biomechanical aspects to the observed kinematics.

2. The tongue model

The high complexity of tongue anatomy makes realistic models of its deformations during speech production very challenging (Perkell, 1974, 1996; Kiritani and Miyawaki, 1975; Kiritani et al., 1976; Kakita and Fujimura, 1977; Kakita et al., 1985; Hashimoto and Suga, 1986; Wilhelms-Tricarico, 1995; Honda, 1996). Indeed, this articulator consists of a complex interweaving of muscular fibers, glands and mucous membranes. Moreover, its shape depends on the recruitment of ten muscles or so, some of them internal to the structure. Thus, just as for the elephant trunk, part of the tongue is responsible for its own deformation. In front of such a complexity,

our modeling approach was not an exhaustive description of the anatomy. Such an objective could only be a long-term project (see Wilhelms-Tricarico, 1995, 1996). Our model is a compromise between resemblance to the reality, design complexity and computation time.

2.1. Soft tissues modeling through the Finite Element Method

The first requirement in building a complete tongue model, is to use an appropriate mathematical tool, to account for soft tongue tissue deformations. The lingual articulator is a continuous and elastic structure, quasi-incompressible since it is mainly composed of water (Smith and Kier, 1989). An efficient mathematical tool to model these properties is the *Finite Element Method* (Schwartz, 1984; Zienkiewicz and Taylor, 1989). The basic aspects of such a modeling are as follows:

- The body is divided into small volumes, the *elements*, which are connected by nodes.
- The discretization of the elasticity equations is made by computing the equations for each node; the trajectories, velocities and accelerations of these nodes are the unknown variables; kinematic properties of the whole body are computed by an interpolation between the nodes.
- Elastic properties of the body are accounted for by two constants: the *Young modulus* E, and the *Poisson ratio* ν (see Appendix A).

Therefore, thanks to this method, it is possible to define, within the elastic structure, different regions (elements or groups of elements), that can have different biomechanical characteristics. This is very convenient to describe the complexity of tongue anatomy, and many authors have adopted this formulation (Kiritani and Miyawaki, 1975; Kiritani et al., 1976; Kakita and Fujimura, 1977; Kakita et al., 1985; Hashimoto and Suga, 1986; Wilhelms-Tricarico, 1995; Honda, 1996).

2.1.1. Geometrical design of the Finite Element structure within the vocal tract

A first important stage in limiting the complexity of the model was to reduce the description of tongue structure to the midsagittal plane. This approach is in accordance with major classical phonetic descriptions that classify speech sounds in terms of high-low/front-back positions of the tongue in the vocal tract. Obviously, some important geometrical features of the tongue such as tongue grooving in [i] or lateral articulations, cannot be accounted for in this way. However, such a description should be relevant enough to assess our model of speech control for the main movements underlying the production of vowels and simple consonants. In addition, this choice is convenient for the evaluation of our hypotheses on speech motor control, since the great majority of tongue movement measurement techniques (microbeam, Electromagnetic Midsagittal Articulometer (EMA), or X-rays) are limited to the midsagittal plane. In the present evaluation procedure, the AG100 EMA system, elaborated by Carstens Electronics in Göttingen, was used. Data and simulations were then compared.

The comparison involves several stages of modeling, each of them being a possible source of distortion. To keep distortions as low as possible, speaker normalization problems were eliminated, by explicitly adapting the geometrical external shape of the model to the tongue contours of a specific speaker. For that, the elements describing the tongue body, were geometrically shaped, to provide a rest configuration to the model. This configuration was extracted from an X-ray picture of the speaker (Badin et al., 1995), close to the production of a *schwa* (see Fig. 1).

Finally the Finite Element (FE) structure was inserted in the vocal tract contours measured on the same X-ray picture (Fig. 1(c)). Muscle insertions on bony parts (jaw and hyoid bone) were simulated by imposing "do not move" constraints to the corresponding nodes. [3] At the bottom of the model (between the mental spine and the hyoid bone), the geniohyoid and mylohyoid effects are functionally

[3] These bony structures are of course likely to move in normal speech. However, to take into account the coupling between their movements and those of the tongue is quite a complex task, involving the interactions of different dynamical models. This was not part of the present work, in which the jaw was blocked and where hyoid bone movements were neglected. This latter assumption is justified by Bothorel's data (Bothorel, 1975) that show that hyoid bone displacements essentially depend on jaw and larynx movements.

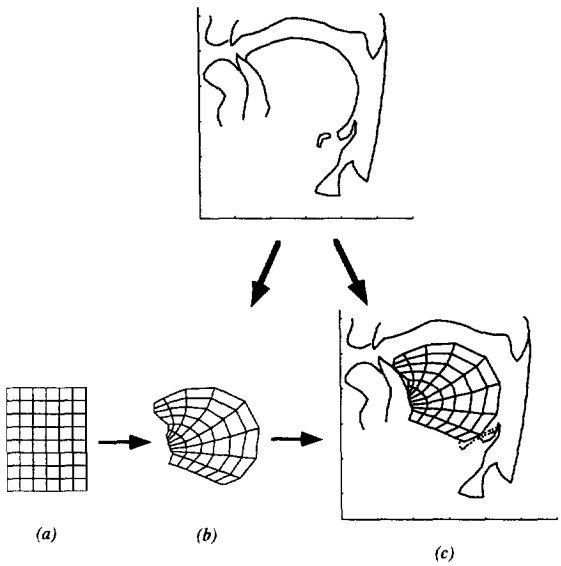

Fig. 1. The 48 elements (63 nodes) building the model (a) are shaped to give the model the rest configuration of the speaker's tongue (b). Finally, this elastic structure is inserted into fixed vocal tract contours (c) measured on the speaker.

modeled as a whole, with a reaction force that is applied to the corresponding nodes to limit the amplitude of any downward movement. As in other biomechanical tongue models published in the literature (Perkell, 1974, 1996; Honda, 1996), the global distribution of the model's nodes reflects the projection of the internal muscle structure into the sagittal plane. The number of elements was determined through the result of a compromise between anatomical accuracy and low computational costs.

2.1.2. Representation of lingual musculature

The great majority of the ten muscles or so that shape the tongue are in fact left–right muscle pairs. Therefore, shaping the tongue is likely to involve the individual control of twenty entities. To our knowledge, there is no evidence of any asymmetrical use of muscles that is relevant for speech communication. Hence, in our modeling approach, limited to the sagittal plane, each muscle pair is modeled as a unique group of fibers. In addition, muscles, whose effects on tongue shaping in the midsagittal plane are slight, are not modeled. For instance, the palatoglossus, which is mainly involved in velum lowering, has not been described. In the same way, the numerous fibers of the transverse muscle have not been represented, their effects being essentially observable in the coronal plane. As for the pharynx constrictors that act on the tongue as well as on the pharyngeal walls, their modeling would imply an additional study to account for the mechanical interaction between these two structures. Thus, they are not part of the current version of the model. Consequently, the muscles described in the model are the anterior and posterior parts of the *genioglossus*, the *styloglossus*, the *hyoglossus*, the *verticalis*, and the superior and inferior parts of the *longitudinalis*.

The mapping between the elements of the structure and the muscles accounts for anatomical data on lingual musculature. This mapping is shown in Fig. 2, in which, for each muscle, the corresponding elements in the model are depicted in gray. The fan-shape of the genioglossus muscle can, for instance, be easily observed inside the FE structure. When a muscle is recruited, the increase of the Young modulus of its associated elements is modeled as a linear function of the level of muscle force. This is in agreement with Duck's (1990) measurements of muscle stiffness during contraction. It should be noted that the mapping is exclusively based on geometrical criteria. Consequently, there is no correlation between the size of an element within the FE structure and the physiological properties of the corresponding muscle. Force generation principles have been modeled separately, as explained in Section 2.2.2. In the model, muscle forces are provided by macro-fibers (bold lines in Fig. 2) that act on selected nodes and simulate the effects of muscles fibers. Their insertion points and directions have been also designed to fit satisfactorily the anatomy. Thus, the two branches of the hyoglossus, namely the basio-glossus and cerato-glossus parts, are represented by two groups of fibers. They are partly external to the tongue, and are inserted into the hyoid bone at their external boundaries. Similarly, the transversal and longitudinal branches of the styloglossus are modeled with two groups of fibers, that are partly external to the tongue. The intrinsic muscle (longitudinalis and verticalis) fibers have been located along the borders of the corresponding elements in the structure. Fig. 2 depicts, in bold lines, the modeled muscular macro-fibers, with their insertions and orientations; the muscular forces are ap-

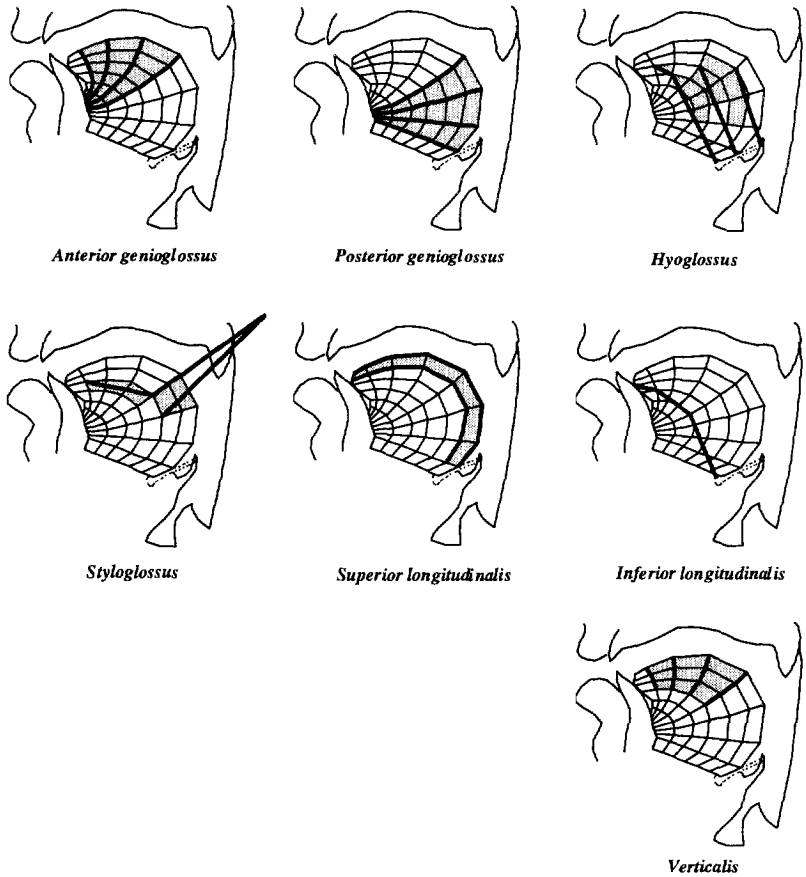

Fig. 2. The macro-fibers, modeling the muscles forces, their orientations, and their insertion points, are shown by the bold lines. The gray-shaded elements are associated with specific muscles; their elastic properties are modified when the muscle is recruited.

plied to the nodes following the direction of the fibers' course.

2.2. Control of the model

2.2.1. The Equilibrium Point Hypothesis

From our point of view, given the discrete structure of the phonological command, a target based model of speech motor control is appealing, as it offers a link between the phonological units and the physical targets that underly the continuous speech signals (MacNeilage, 1970). From this perspective, Feldman's EPH (Feldman, 1966, 1986) is highly interesting for speech production, as its theoretical framework is related to the notion of target.

This theory of motor control suggests that movements are produced from posture to posture, and it functionally accounts for basic neurophysiological mechanisms of muscle force generation. From experimental studies of torque variations in forearm movements, Feldman (1966) has suggested that the Central Nervous System (CNS) would control movements by changing, for each active muscle, the threshold muscle length, λ, where the recruitment of α motoneurons (MNs) (responsible for active forces) starts. The EPH assumes that afferent inputs related to muscle length and velocity are acting, together with descending central input, on the α MNs, to produce a global level of muscle force. Note that this theory was originally proposed for skeletal muscles, and that the feedback was assumed to be due to muscle spindle activity. It was extended to the jaw/hyoid bone set (Flanagan et al., 1990; Laboissière et al., 1996), and to the oculomotor

system (Feldman, 1981). Some tongue muscles, in their structure, are different from skeletal muscles; in particular, to our knowledge, it was never demonstrated that the muscle spindles located in tongue muscles (Cooper, 1953; Walker and Rajagopal, 1959) are used for postural control. However, proprioceptive feedback has been observed in the control of tongue position (Weddell et al., 1940; Adatia and Gehring, 1971). This suggests that, in the tongue also, some mechanoreceptors may provide the afferent facilitation to the α MNs.

According to the EPH, if muscle length is longer than λ, muscle force increases with the difference between the two lengths, according to an exponential force-length relationship called *Invariant Characteristics*. Thus, for a given muscle length, the muscle force depends on λ value. A posture being a stable mechanical equilibrium state of the motor system, it is fully determined by the λ values of the active muscles. Hence, according to Feldman, the continuous trajectory of the end-effector would be centrally controlled by a discrete sequence of λ *control vectors* that specify the successive postures. Feldman's EPH is thus well adapted to be used in the framework of a target-based model. In addition, for a constant amount of external forces, a mechanical equilibrium of the motor system is associated with a specific spatial position of the articulator. Hence this model also proposes a way to relate motor control vectors to the physical space in which the movement is produced.

Another important assumption of the EPH is that movements are produced with simple time variations of the control vectors. This determines a *virtual trajectory* in λ space (see Hogan, 1984, for a presentation of the notion of virtual-trajectory), and the differences between virtual and actual trajectories depend on the dynamical properties of the system. More specifically, Ostry and his colleagues (Feldman et al., 1990; Flanagan et al., 1990; Ostry et al., 1992) have proposed that control vectors should vary at a constant rate shift from posture to posture.

Thus, the basic principles for the control of the tongue model are as follows (Payan et al., 1995):

- Movements are produced towards spatial equilibrium configurations.
- Between two targets, control vectors are shifted at a constant rate.
- Proximity between actual (physical space) and virtual (λ space) trajectories can be adjusted by tuning the dynamical parameters of the system as well as the timing (transition and hold durations) of the commands.

We propose that the equilibrium target associated with a phoneme is unique for a given phonetic context. This target is thus assumed to be independent of prosodic commands such as speaking rate or emphasis. However, for the same phoneme, this target should a priori change, if the phonetic context varies, as a result of a planning process. We developed elsewhere (Perrier et al., 1996a) how such a process could specify spatial targets as a function of the phonetic context (see also Guenther (1995) for similar concepts related to orosensory targets).

2.2.2. Force generation modeling

The modeling of tongue muscle force generation has been directly inspired from the work carried out by Feldman and his colleagues (Feldman, 1966, 1986; Laboissière et al., 1996).

To account for the observed exponential shape of the Invariant Characteristics (Feldman and Orlovsky, 1972), the muscle force $\tilde{M}$ is functionally described as an exponential function of the muscular activation A. Thus,

$$\tilde{M} = \rho[\exp(cA) - 1], \tag{1}$$

where c is a form parameter, and ρ a magnitude parameter directly related to force-generating capability. In addition, it should be noted that a saturation is applied when $\tilde{M}$ reaches its maximal value ρ. Parameter c has been estimated from force-length relations reported by Feldman and Orlovsky (1972) and empirically obtained for a cat gastrocnemius muscle. In our simulation experiments, its value has been fixed, for all muscles, to $0.112\,\mathrm{mm}^{-1}$. The ρ parameter depends conversely, on each specific muscle force-generation capability, and cannot be chosen as a constant value for all muscles. The ρ values were obtained by estimating the muscle cross-sectional areas, as suggested by Laboissière et al. (1996). For that, data available in anatomic atlases (Gerhardt and Frommhold, 1988; Gambarelli et al., 1977) were used, as well as detailed drawings of tongue cross-sections provided by Miyawaki (1974). The ρ values used in the model are given in Table 1.

Table 1
Cross-sectional area of the muscles and the corresponding ρ values in the model

Muscle	Cross-sectional area (mm^2)	$\rho(N)$
Genioglossus	308	67
Hyoglossus	296	65
Styloglossus	110	24
Superior longitudinalis	88	19
Inferior longitudinalis	65	14
Verticalis	66	14

The muscle activation A is represented by the difference between the actual muscle length l and the dynamical recruitment threshold λ'. λ' is a function of the recruitment threshold λ and the muscle shortening or lengthening velocity $\dot{l}$. Thus,

$$A = [l - \lambda']^+ = [l - \lambda + \mu\dot{l}]^+, \quad (2)$$

with, by convention,

$$[x]^+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x \leqslant 0, \end{cases} \quad (3)$$

where μ is a damping coefficient due to proprioceptive feedback. Its value was fixed to 0.05 second, for all muscles (Laboissière et al., 1996).

Thus, the muscle force is computed as the result of the interaction between a descending central input (the λ command) and afferent inputs related to muscle length and velocity.

In addition, the *sliding filament theory* (Huxley, 1957) was taken into account in the model: the force generation capability of a muscle depends on muscle shortening or lengthening velocity. The heads of the myosin filaments (the *cross-bridges*) are elastic and can interact with a given actin site over a relatively wide range of distances. To model this dependence, we have fitted the force-velocity relation measured by Wells (1965), for the anterior tibial cat muscle. Because of absence of data on the tongue, Wells' measurements were chosen, because they were obtained for a *fast* muscle, and the assumption has been made that tongue muscles are *fast* muscles. According to Laboissière et al. (1996), these data were fitted with a sigmoid curve. The actual active force F is then computed as follows:

$$F = \tilde{M}\left[f_1 + f_2 \operatorname{atan}\left(f_3 + f_4 \dot{l}/r\right) + f_5 \dot{l}/r\right], \quad (4)$$

with $f_1 = 0.80$, $f_2 = 0.50$, $f_3 = 0.43$, $f_4 = 0.80$ and $f_5 = 0.023$. r is the length of the muscle when the model is in its rest configuration (see Fig. 1), l the muscle length and $\dot{l}$ its first time derivative.

2.3. Modeling tongue deformations over the time

2.3.1. Elasticity parameterization

In designing the model, an important stage consisted in determining the constants of elasticity, namely the Young modulus E and Poisson's ratio ν. Indeed, their values strongly influence the way each muscle can contribute to the global shaping of the tongue.

Poisson ratio specifies how a force applied on the structure in a given direction will induce deformations in the other directions (see Appendix A for details). In the model, this parameter was set to 0.49, to account for the quasi-incompressibility of tongue tissues.

The Young modulus influences the deformation of the structure in the direction of the force; it basically measures the stiffness of the material (see Appendix A for details). Unfortunately, no data is available in the literature about this constant for tongue tissues. Nevertheless, measurements are reported for other parts of the body. For the vocal folds, Min et al. (1994) have measured a Young modulus of 20 kPa, while Oka (1974) has reported a value of 300 kPa for a contracted skeletal muscle. Duck (1990) found a value of 6.2 kPa for a human muscle in its rest position, and a value of 110 kPa for the same muscle when it was contracted. Hence it seems that a positive correlation exists between muscle recruitment and Young modulus.

In the current version of the model, a value of 15 kPa was chosen for the elements corresponding to passive tissues. As for the elements corresponding to active muscles, it was arbitrarily proposed to model the positive correlation by a linear relationship. The Young modulus increases linearly with the muscle force F, from 15 kPa (no activation, $l \leqslant \lambda$) to a maximum of 250 kPa (maximal value of $(l - \lambda)$). These boundary values are consistent with the measurements reported just above. They were empirically found, from several simulations with the model, to be able to generate muscle force levels compatible with those measured during vowel gestures by Bun-

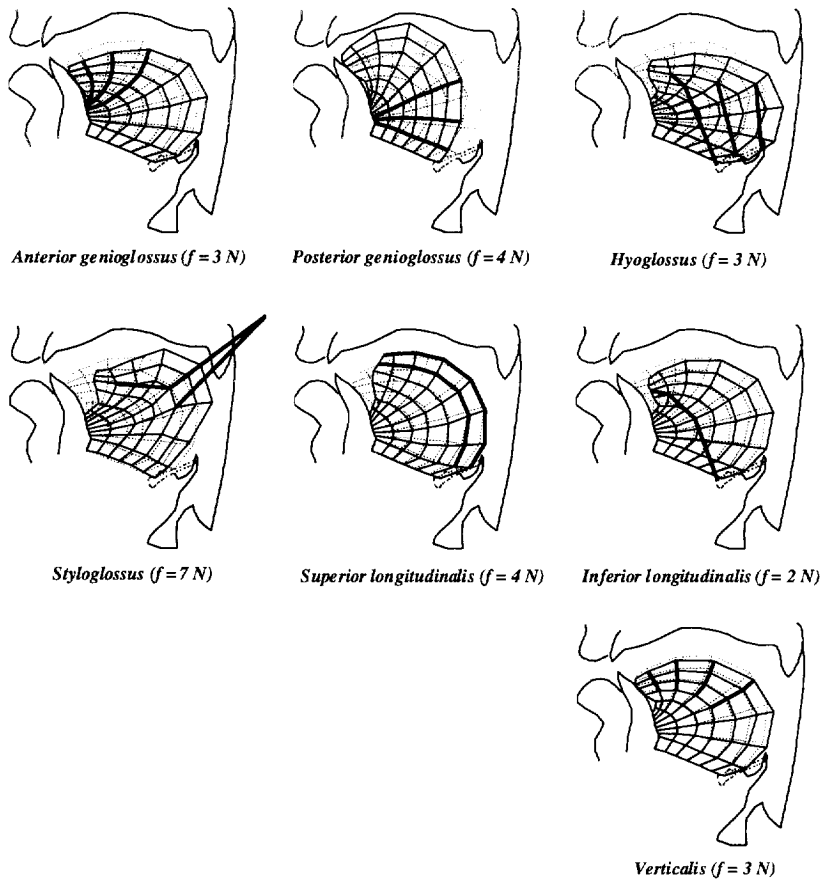

Fig. 3. Contribution of each muscle to the global tongue shaping. The levels of the generated forces (in Newtons) are given at the bottom of each panel. The dotted mesh shows the biomechanical structure in its rest configuration.

ton and Weismer (1994) (see also Dworkin et al., 1980).

With these values for the constants of elasticity, the structure of the 2D tongue model is completely defined. Fig. 3 shows, for our model, the contribution of each muscle to the global shaping of the tongue.

2.3.2. Equations of motion

Let U, $\dot{U}$ and $\ddot{U}$ be respectively the displacement, velocity and acceleration vectors (dimension: 63×2) of the structure's nodes. The temporal evolution of each of these vectors is described by the classical equation of motion:

$$M\ddot{U} + f\dot{U} + K(U,\dot{U},\lambda)U = F(U,\dot{U},\lambda) + P. \qquad (5)$$

M is the global mass matrix (100 g for the whole tongue, each node mass being proportional to the area of the corresponding element); f is the global passive damping matrix, with a coefficient of viscosity set to 15 N s/m; F is the active muscle force, and P represents the external forces, namely the gravity and the contact reaction forces due to interactions between tongue and palate or tongue and teeth. K is the stiffness matrix which determines the internal forces within the FE structure, and accounts for the distribution stiffness over the elements (as imposed by the association between muscles and elements in the model). This matrix is computed by the FE algorithm.

The solving of this equation, based on an adaptive Runge Kutta method, gives the movements of each

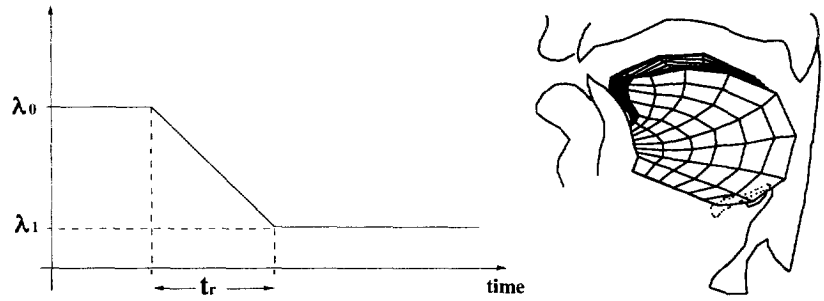

Fig. 4. Tongue movement produced by a constant rate shift of the recruitment threshold of the hyoglossus, no other muscle being recruited.

node over the time, and then the temporal evolution of the tongue shape. Fig. 4 shows an example of a movement produced with a linear shift of the hyoglossus recruitment threshold.

3. Simulations with the model

As mentioned above, the evaluation of our model of speech movement control lies in a comparison between simulations and data collected on a human speaker. In this part, the corpus and the experimental protocol will first be presented. Secondly, some simulation results will be shown, and finally an analysis of the potential influence of biomechanical properties of the articulators on kinematic features will be carried out.

3.1. Reference data acquisition

3.1.1. Experimental protocol

The corpus was designed to study V–V sequences, where tongue is the most relevant articulator. Indeed, assessing whether the observed articulatory and acoustic signals essentially depend on tongue movements, is a prerequisite, if our simulations are to be a reliable evaluation of the control model. Hence the corpus consisted of sequences where lip movements are acoustically not relevant, and a cubic bite-block was inserted between the teeth of the subject, to keep his jaw position constant. The size of the bite-block (6 mm) was chosen by the subject himself, to make sure that no inconvenient perturbation of the speech production was induced. Preliminary acoustic recordings were carried out, to verify that the bite-block did not actually induce an observable reorganization of the speech production strategy. The acoustic signals recorded under normal and bite-block conditions did not present any significant spectral differences. The formant patterns for each vowel, and the shapes of the transitions were essentially identical; the only difference consisted of a slight decrease in speaking rate for the bite-block condition.

During the experiment, the subject produced all possible V–V sequences within two sets of vowels: the spread vowels [i, e, ε, a] and the rounded ones [y, o, u]. The subject was asked to produce each sequence with as little change in his labial shape as possible, and to hold each successive vowel for a sufficient amount of time, to avoid target undershoot.

Articulatory data were collected with a five-transducers Electromagnetic Mid-sagittal Articulometer (AG100 system by Carstens Electronics). Three transducers were glued onto the tongue. They were regularly located, from 1 cm through 5 cm from the tongue tip. The remaining transducers were glued onto the superior and inferior incisors. All of them were located as accurately as possible in the mid-sagittal plane.

In addition, front views of the lips were recorded with a video camera. For later automatic extraction of lip contours, involving a Chromakey processing, the lips of the subject were carefully painted in blue (Lallouache, 1990). These video recordings have been essentially used to check whether lip movements were effectively negligible within each sequence. The mean lip area was also calculated for each sequence to serve as input for the synthesis of the acoustical signal associated with the simulated tongue movements.

3.1.2. The subject

The subject was a male native speaker of French, PB, who volunteered for this experiment. He is one of the "reference speakers" at the *Institut de la Communication Parlée*. Therefore, he is used to this kind of experiment. Cineradiographies and corresponding speech signals are already available for this subject (Badin et al., 1995), and they were used for the current work, even if they were collected for other purposes and other corpora. X-ray pictures provided information about the whole geometry of the vocal tract in the mid-sagittal plane, complementary to the data obtained with EMA. A model of the transition from the mid-sagittal contours to the area function of the vocal tract was also used in the simulation procedure. It was specifically built for this subject from the correspondence between formant values and X-ray data (Beautemps et al., 1995).

3.2. Controlling the model for Vowel–Vowel sequences

To simulate Vowel–Vowel (V–V) sequences, a set of motor commands has to be associated with each vowel, and a temporal variation of these commands, from one vowel to the other, has to be proposed. For that, we have used additional simplifications that can be justified for vowel production.

3.2.1. Mapping between tongue geometry and recruitment thresholds

In theory, because of the numerous compensation possibilities between muscles, there is an infinity of sets of recruitment thresholds that could correspond to a same tongue shape. However, in most cases, from one of these sets to the other, the global level of muscle force may vary. Therefore, a way to minimize the number of possibilities is to impose a certain level of muscle force. The mapping will even become unique, if each muscle force is constrained to be equal to zero. Such a configuration is theoretically possible: it corresponds to the case where the actual lengths of the active muscles are equal to their recruitment thresholds. However, in reality, whatever the tongue shape, the external gravity force would tend to pull the tongue down, while passive elastic reactions would pull the tongue back to its rest-configuration. Hence, for each tongue posture, a minimum level of muscle force is always required to just counteract those external loads. For that, the recruitment thresholds for the active muscles can never be equal to the actual muscle lengths; they must be slightly smaller.

In this paper all V–V sequences have been generated by producing the target tongue shapes of the vowels with the minimum level of force. This provides a simple way to propose a mapping between the recruitment thresholds of active muscles and the external contours of the tongue. The basic principle underlying this approach consisted, for each target tongue shape, of the following stages:

- Defining which muscles are active.
- Measuring the length of these active muscles in this tongue configuration.
- Finding the desired recruitment thresholds of the active muscles, by starting from their actual length and, then, by slowly and manually decreasing their values down to the point where the muscle forces are just sufficient to counteract the external forces and to hold the posture.

To identify the active muscles, a simplification was provided by observations made by Maeda and Honda (1994) (see also Honda, 1996) on tongue EMG data. These authors showed that, for a constant position of the jaw, the tongue shapes observed during vowel production can be fairly well explained by the recruitment of only three muscles: the genioglossus, divided into its posterior and anterior parts (GGp and GGa), the styloglossus (SG) and the hyoglossus (HG). Consequently, in our model, the verticalis and the superior and inferior longitudinalis were assumed not to be active during V–V sequences.

Data provided by Honda and colleagues (Baer et al., 1988) give more details about the amplitudes of the EMG signals for vowels of American English. It is thus possible to identify which is the most active among the three above mentioned muscles: GGp for [i], HG for [a] and SG for [u]. In the present simulations, only these three main muscles were likely to be active. Thus, in the first assessment of the model, we have accounted for a simplified view of Honda's (1996) suggestions, and have generated Vowel-to-Vowel transitions from simple underlying schemata of muscle activity:

- To move the tongue from a front-high to a back-low position (the [i-a] direction), and reciprocally, the only active muscles are GGp and HG.

- To move the tongue from a front-high to a back-high position (the [i-u] direction), and reciprocally, the only active muscles are GGp and SG.
- To move the tongue from a back-high to a back-low position (the [u-a] direction), and reciprocally, the only active muscles are SG and HG.
- To move the tongue in any combination of these three basic directions, more than two muscles are recruited; for instance in the [y-o] transition, GGp, SG and HG are assumed to be active.

In this framework, the recruitment thresholds associated with the minimum level of force were determined as follows, for the extreme vowels [i], [u], [a]:

- To hold the posture of vowel [i], the only active muscle was assumed to be GGp. Its recruitment threshold was determined as described above; as for the other muscles, because they are not active, any λ value can be used, as long as it is larger than the actual muscle length in the posture (no generation of active forces). However, according to Honda, HG is active during [i-a] transition. We have therefore adjusted its recruitment threshold for vowel [i] to its length in this tongue posture, so that it would be recruited as soon as the movement toward vowel [a] begins. For the same reason, in the context of the [i-u] transition, the recruitment threshold of SG was constrained for [i] to be equal to its actual length.
- For vowel [a], similar principles were applied; the only active muscle was assumed to be HG; and it was supposed that the recruitment thresholds of GGp and SG are equal to their actual lengths in the [a] tongue posture. GGp and SG were respectively recruited during [a-i] and [a-u] transitions.
- For vowel [u], the same principles were applied; the only active muscle was assumed to be SG; and it was supposed that the recruitment thresholds of GGp and HG are equal to their actual lengths in the [u] tongue posture. GGp and HG would be respectively recruited during [u-i] and [u-a] transitions.

Clearly, these patterns are too simplistic to describe, in its entirety, the complexity of muscle activation during speech production. However, we think that they account for the main features of muscle activation. Hence, as the main objective of this study is to operate the tongue model with our hypotheses for its control, and to limit their assessment to a qualitative comparison of recorded speech data and synthesized sequences, we believe that it is not necessary to account for the whole complexity.

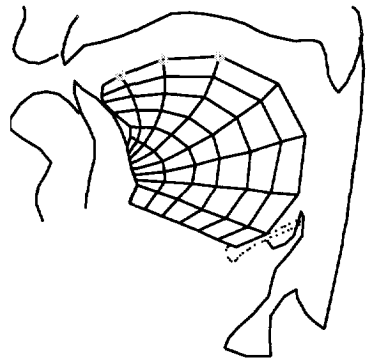

Fig. 5. The three nodes of the FE structure associated with the three transducers glued onto the speaker's tongue.

To find geometry of the tongue model associated with each of these three vowels, two sets of articulatory data, collected on the same speaker, were used: the EMA recordings described above, together with cineradiographic pictures, collected for another corpus and other purposes (Badin et al., 1995). Within the FE structure, three nodes, located on the tongue surface in the palatal part, were selected. These nodes represent the three EMA transducers glued onto the speaker's tongue. (See Fig. 5.) For each of the vowels, the model was then shaped manually, by acting on the corresponding active muscles, to get a satisfactory fit between the position of the three nodes and that of the transducers. As only one or two muscles are active (see above), it is not possible to obtain a perfect fit. Hence the discrepancy between EMA data and node positions could be equal for several shapes. In arriving at a decision, the information available from the cineradiographic data was helpful since the X-ray images show the whole sagittal contour. However, as the experimental protocol used for these data did not include any bite-block constraint, only pictures showing a jaw position similar to the one imposed by the bite-block in our EMA experiment were useful. Thus, these data were considered only for the closed vowels [i] and [u], and the tongue shapes which resembled most faithfully the X-ray contours, were selected. As regards [a], received knowledge about the tongue shape was used to make a decision.

For the other vowels included in the reference EMA corpus, [y], [e], [ε] and [o], the motor commands were found as follows:

- [y] was assumed to feature essentially the same tongue posture, and the same motor commands for the tongue, as [i]; hence it was supposed that the [i] versus [y] distinction is mainly due to lip gesture;
- [e] and [ε] were considered as intermediate vowels on the trajectory between [i] and [a]; hence it was assumed that their motor commands are intermediate between those of [i] and [a], following a linear function:

$$\lambda_{GGp}([e]) = \lambda_{GGp}[i] - k_{[e]}(\lambda_{GGp}[i] - \lambda_{GGp}[a]), \tag{6}$$

$$\lambda_{HG}([e]) = \lambda_{HG}[i] - k_{[e]}(\lambda_{HG}[i] - \lambda_{HG}[a]) \tag{7}$$

and

$$\lambda_{GGp}([\varepsilon]) = \lambda_{GGp}[i] - k_{[\varepsilon]}(\lambda_{GGp}[i] - \lambda_{GGp}[a]), \tag{8}$$

$$\lambda_{HG}([\varepsilon]) = \lambda_{HG}[i] - k_{[\varepsilon]}(\lambda_{HG}[i] - \lambda_{HG}[a]). \tag{9}$$

Parameters $k_{[e]}$ and $k_{[\varepsilon]}$ are ad-hoc values, that provide a good fit between the three reference nodes in the model and the positions of the three transducers.

- SG and HG are the active muscles which hold the posture of vowel [o]; in the context of the [o-i] transition, the recruitment threshold of GGp is constrained for [o] to be equal to its actual length.

3.2.2. *Generation of V–V sequences*

From the above model of tongue control, the principles underlying the synthesis of V–V sequences can be seen to be fairly straightforward. Tongue movements were generated by a constant rate shift of the recruitment thresholds of the active muscles, between the threshold values for the two targets of the sequence; the onset and offset times were identical for all motor commands shifted during the sequence, and they were manually tuned to correctly approximate the duration of movement measured on the speaker's data. Fig. 6 gives an example of this command pattern for the [i-a] sequence.

Since it was possible to verify from the video recordings that, within each sequence of the EMA corpus, lip movements were slight, lip area was kept constant throughout the synthesis of a V–V transition, and was equal to the area measured for one of the vowels of the sequence. Similarly, the vertical

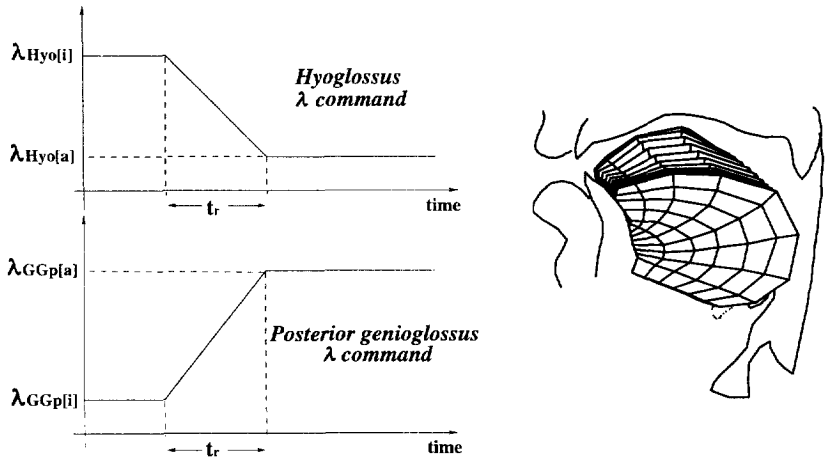

Fig. 6. Generation of the [i-a] sequence with the tongue model: temporal evolution of the motor commands (left) and tongue movement (right). According to Eq. (2), muscle activity increases as the recruitment threshold λ decreases. GGa and SG are presumed to be inactive.

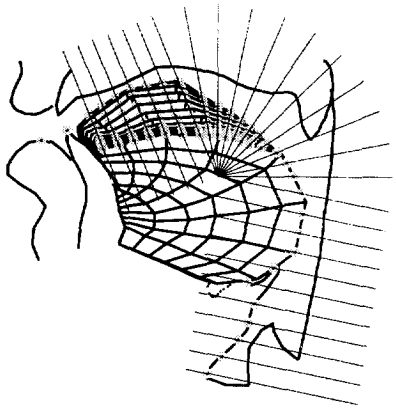

Fig. 7. The measurement of the vocal-tract edges was made through the intersection of the external contours of the model and the fixed walls, with a grid plotted in the sagittal plane.

movements of the larynx were neglected, and the hyoid bone was considered as a fixed structure.

To generate the acoustic signal from the sequences of tongue shape produced by the model, a computation of the area function was achieved. Following a classical approach, the external contours of the model and the fixed pharyngeal walls were projected onto a measurement grid in the sagittal plane (Maeda, 1979). A discretization of these contours was made through their intersection with the grid (Fig. 7). This provided a function describing the vocal-tract edges in the mid-sagittal plane, from which the area function was calculated following the procedure [4] proposed by Beautemps et al. (1995). The speech signal synthesis was then based on Pelorson et al.'s acoustical model (Pelorson et al., 1995). In parallel, the transfer function was computed with a harmonic model (Badin and Fant, 1984), and the first three formants were calculated.

Several V–V sequences were simulated. Including all of them in this paper would be too long. Hence the presentation of the results will only focus on three examples, the [i-a], [y-u] and [y-o] transitions, that exemplify the ability of the model to produce realistic articulatory and acoustic signals, as well as the influence of biomechanical properties of the motor system on the kinematic features.

3.2.3. Evaluation of synthesized acoustic signals

The assessment of the speech synthesis was not based on any quantitative comparison at the acoustic level. Indeed, as explained above, the speech synthesis involves a succession of models, from the sagittal contours of the vocal tract to the acoustics. Even though each of these models has been carefully elaborated, there is at each stage the risk of a slight discrepancy between the model and the speaker data. Moreover, the geometry of the model is sometimes inaccurate, as for instance for [a] and [o], where no information about the pharyngeal part was available. Similarly, the movements of the hyoid bone due to laryngeal displacements were neglected, while lip area was kept constant during the transitions. Hence the computation of distances between the signals in the temporal or spectral domain could reveal large errors, even though both signals were qualitatively very similar. For this reason, the assessment of the synthetic signal consisted in evaluating the global similarities in the shape of the formant trajectories.

Fig. 8 shows the natural and the simulated formant trajectories in the F_1/F_2 plane, for [i-a] and [y-u]. It can be observed that the simulated trajectory of the [i-a] sequence, like the measured data, is essentially straight, keeping constant its mean direction, from [i] to [a] via the vowels [e] and [ɛ]. The main feature of the natural [y-u] transition is that formant F_1 remains fairly constant during the transition. This characteristic requires accurate articulatory coordination between the horizontal and vertical displacement of the tongue. The backward movement of the tongue from [y] to [u] reduces the volume of the vocal tract back cavity; this reduction should induce an increase in the Helmholtz resonance [5] F_1. To keep this value constant, the volume decrease should be balanced by maintaining a high tongue position throughout the transition. With a geometrical model

[4] It should be noted that these authors have elaborated this procedure from the cineradiographic data that we have used afterwards in defining the mapping between tongue shapes and motor commands.

[5] In the French vowels [y] and [u], F_1 is the Helmholtz resonance of the set "back cavity + constriction" (Fant, 1960; Badin et al., 1990): $F_1 = (c/2\pi)\sqrt{A/lV}$, where A and l are respectively the area and the length of the constriction, V is the volume of the back cavity, and c the sound propagation velocity.

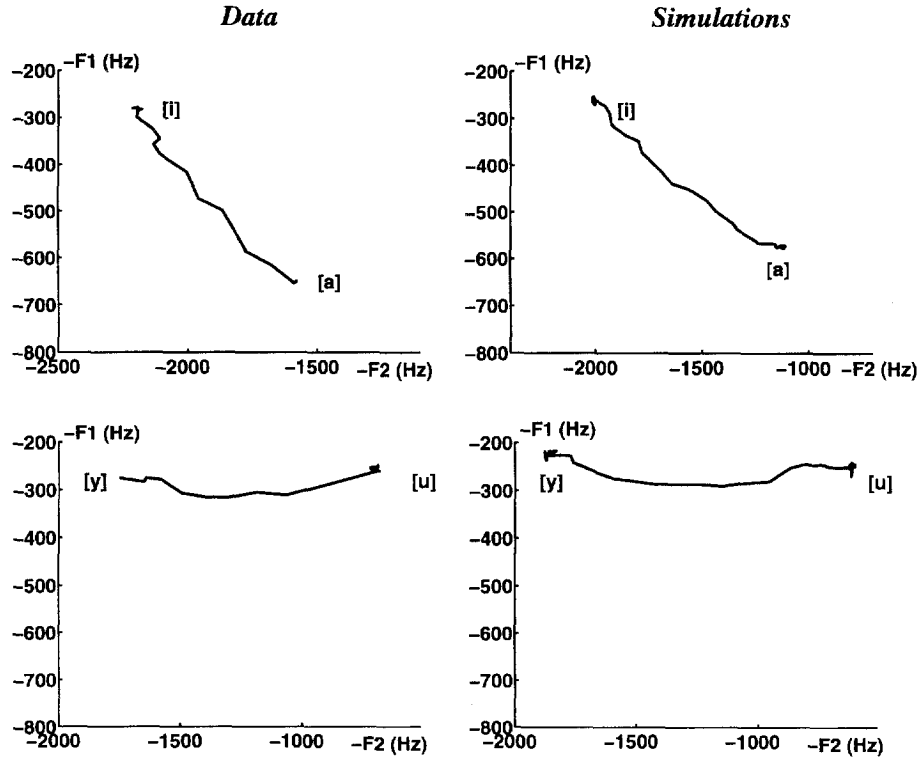

Fig. 8. Formant trajectories in the F_1/F_2 plane: data measured on speaker PB are displayed on the left, and the simulations on the right. Upper panel: [i-a]; lower panel: [y-u].

of the vocal tract (e.g. Fant, 1960, 1992) such a constraint has to be explicitly specified (for instance, by carefully controlling the size of the constriction from [y] to [u]). Carré and Mrayati (1991) have presented results exemplifying the consequences of the absence of such a constraint. In our anthropomorphic tongue model, this balance is implicitly simulated through a realistic anatomical description and an account of tongue shaping.

3.2.4. Evaluation of the articulatory patterns

In the same way as for the acoustic signals, the assessment of the articulatory patterns was not based on a quantitative comparison between simulations and data. The proximity between node and transducer trajectories was thus only qualitatively evaluated. Articulatory velocity profiles, depicting the time variation of the velocity, were therefore observed. This was justified by the fact that many studies of human motor control are based on the analysis of the kinematic variable (Morasso, 1981; Soechting and Lacquaniti, 1981; Nelson, 1983; Ostry and Munhall, 1985; Munhall et al., 1985; Ostry and Flanagan, 1989; Adams et al., 1993). To illustrate how biomechanical models can be helpful in studying human motor control, only [y-o] and [y-u] transitions are described. Fig. 9 (respectively Fig. 10) shows the measured data and the results of the simulations for the [y-o] (respectively [y-u]) sequence.

Each figure is divided into two panels. The upper panel, in its upper part, depicts the back transducer velocity profile for three repetitions of the sequence, and, in its lower part, the simulated velocity profile of the corresponding node in the FE structure. The lower panel shows the same results for the middle transducer. For both sequences and both transducers, a good similarity can be observed between data and simulations. However, the simulated profiles are more skewed, as the initial slope seems too large. This weakness of the model originates from a crude

estimation of mechanical parameters such as tongue mass, viscosity, ρ and μ parameters.

A good feature of the model is that it is able to predict the differences noted in the data between the [y-o] and the [y-u] velocity profiles. Whereas only one velocity peak is observed for the [y-o] sequence (in Fig. 9), two peaks are noticed for [y-u] (in Fig. 10). This discrepancy could be interpreted as a specificity in the motor control of vowel [u] as compared with vowel [o]. This would be consistent with Baer et al.'s (1988) data for English vowels. These authors have shown that vowel [u] is associated with an

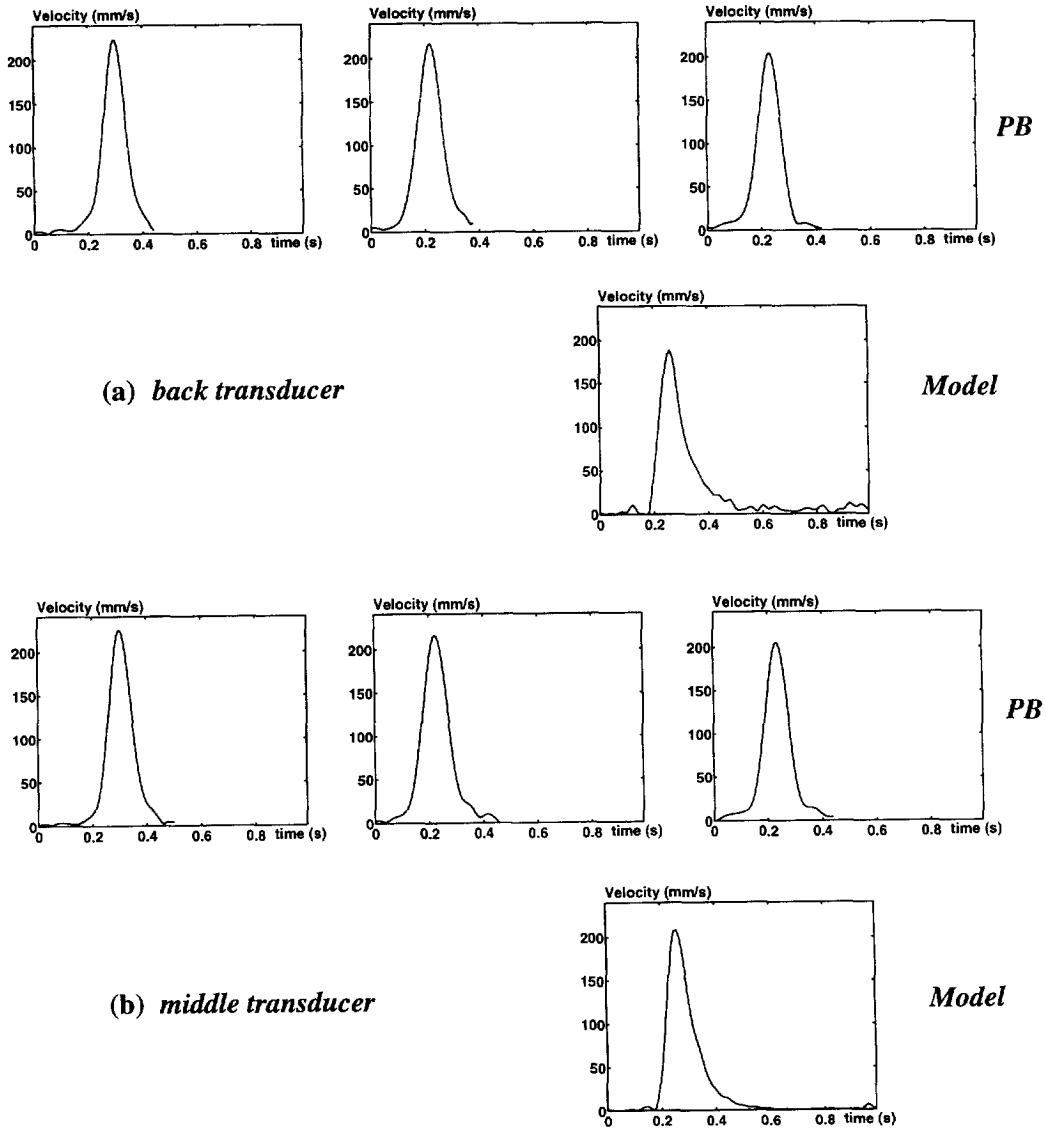

Fig. 9. Velocity profiles of the [y-o] sequence as generated by the model and as exemplified through three repetitions by speaker PB. Upper panel: back transducer (a); lower panel: middle transducer (b).

activation of SG followed by an activation of HG. Thus, one could speculate that SG activation produces the first peak, while the second peak is due to HG. This phenomenon could be the result of a diphthongization of vowel [u]. However, French vowels are usually not diphthongized. Moreover, in our data, formant patterns are stable during the whole vowel, so that this hypothesis can definitely be excluded. Hence an explanation of this phenomenon which is consistent with the hypothesis of a simple articulatory gesture, is to be preferred. In our tongue model, this complex velocity pattern can be explained with a simple target-to-target transition. This is demonstrated below.

In the simulation of the [y-o] sequence, SG and HG are the agonist muscles. They cooperate to move

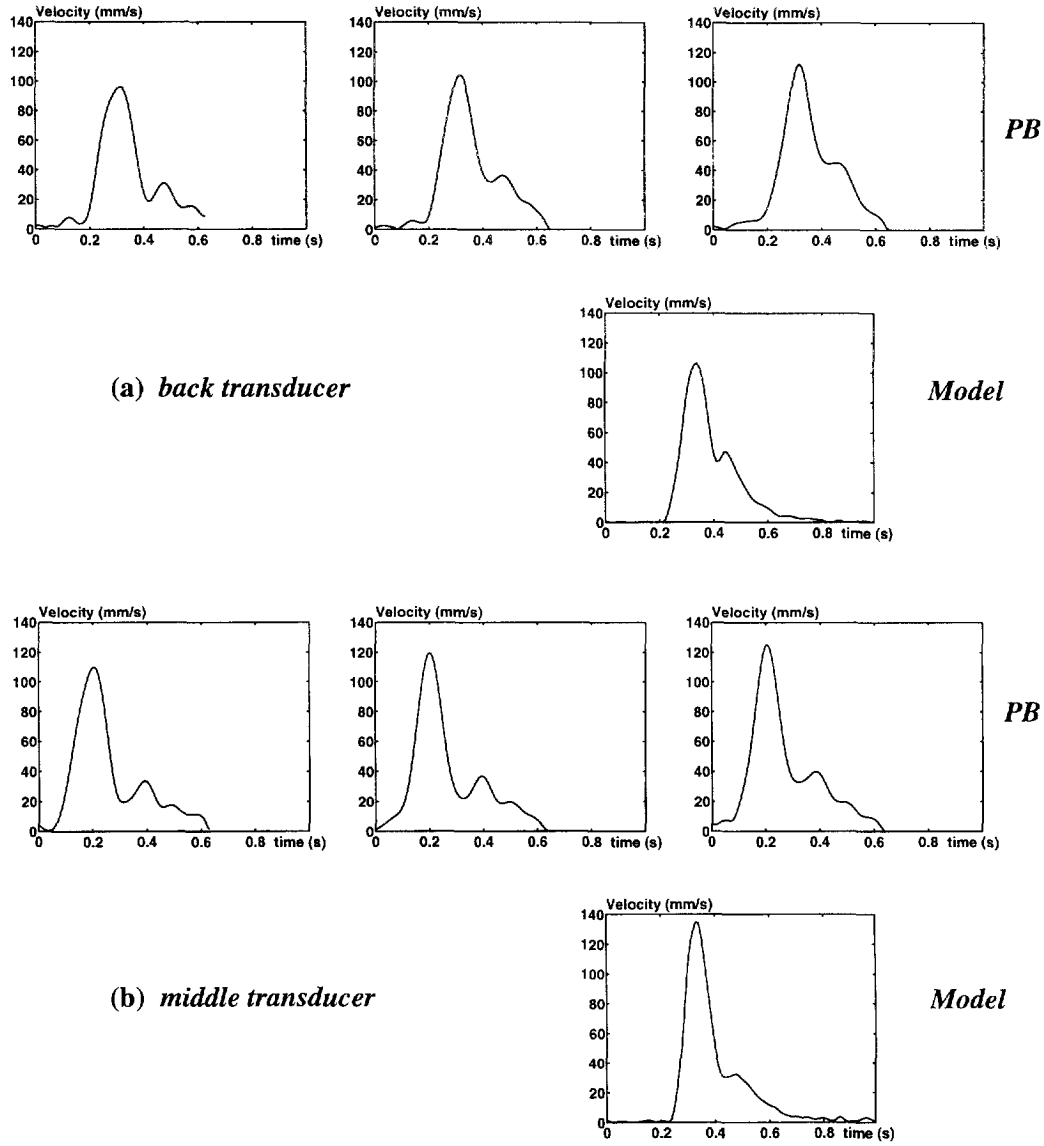

Fig. 10. Velocity profiles of the [y-u] sequence as generated by the model and as exemplified through three repetitions by speaker PB. Upper panel: back transducer (a); lower panel: middle transducer (b).

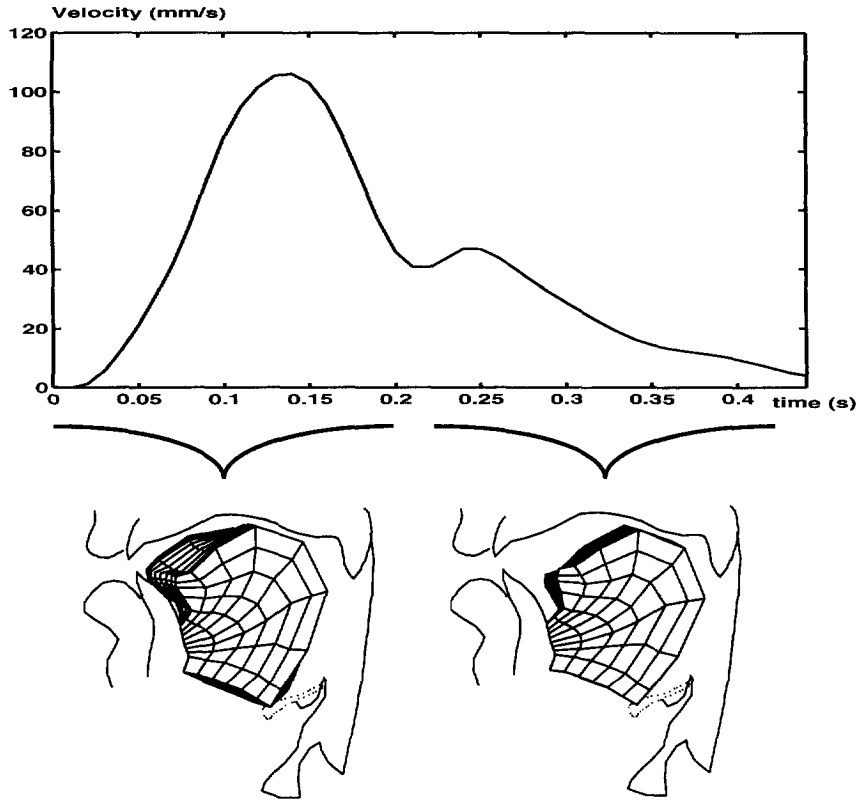

Fig. 11. Correspondence between the two parts of the velocity profile of the back node (top panel) and tongue displacements (low panel) in the [y-u] transition generated with the model.

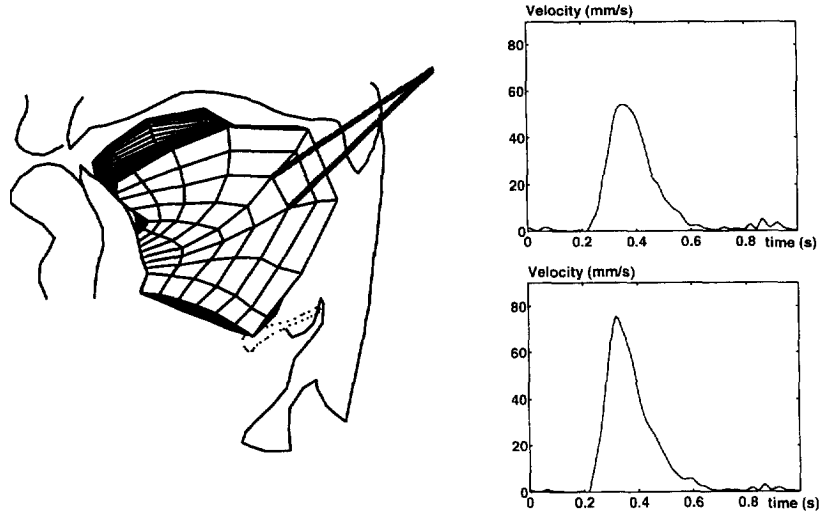

Fig. 12. Tongue displacements (left panel) and the corresponding velocity profiles of the back node (right panels) in the [y-u] transition generated with a modified version of the model (after removing of the anterior part of SG fibers).

the tongue backwards. For the [y-u] sequence, SG is the only agonist of the movement; as shown in Figs. 2 and 3, the direction of the muscle's fibers changes within the tongue. In their front part, fibers are essentially oriented along the horizontal axis, whereas in their rear part they are oriented towards the styloid process. Our suggestion is that this morphological property is responsible for the double peak. This analysis is illustrated in Figs. 11 and 12.

Fig. 11 shows the velocity profile (top panel) and the corresponding sagittal contours of the model (low panel), for the synthesized [y-u] transition. The sagittal views are depicted in two sets: the left one corresponds to the movement from time 0 (onset of the gesture) to the time where the velocity reaches its minimum between the two peaks; the remaining sagittal views, up to the end of the movement, are shown in the right set.

The data presented in Fig. 12 were generated with a modified version of the model: the front part of the SG fibers has been removed (see left panel). These data again correspond to a simulation of the [y-u] transition, still with a constant rate shift of the control vector. The associated velocity profiles of the two tongue surface nodes are plotted in the right panels. It is interesting to note that these profiles depict a single peak. Thus, removing the front part of the SG fibers has a strong effect on the kinematic features of the movement. This can be explained by comparing the respective sagittal views of the two transitions (Fig. 11, low panel versus Fig. 12, left panel). Without the front SG fibers (Fig. 12, left panel) the anterior part of the tongue first moves back under the coordinated actions of GGp deactivation and SG activation; once SG effect begins to dominate and raises the central part of the tongue, the backward movement slows down. With the complete model (Fig. 11, low panel), we can observe similar changes in the first part of the movement. However, once SG effect begins to dominate, the raising movement is associated with a bunching action, due to the contraction of the front fibers. It provokes a slight second backward movement of the anterior part of the tongue.

These results emphasize how the morphology of the motor system can influence the kinematics. They are relevant in the studies of motor control strategies. Several works have considered the number of velocity peaks in order to infer the underlying control strategies. For instance Adams et al. (1993) have proposed that multiple submovements, using afferent information, would explain the kinematic features observed in tongue tip lowering gestures. Another example can be found in Boyce et al. (1990) who have suggested that the two velocity peaks measured in a lip protrusion gesture would correspond to a two-stage articulatory pattern. From our point of view, such conclusions should be considered with caution. The study carried out by McClean and Clay (1995) suggests that the variability in the number of velocity peaks, observed across speaking rates, could have neurophysiological explanations, related to the firing rate of motor units. In other respects, our simulations demonstrate that kinematic studies should consider biomechanical aspects of the motor system, to assess the validity of their conclusions.

4. Conclusion

Feldman's Equilibrium Point Hypothesis was applied to the control of a 2D biomechanical model of the tongue for the production of V–V sequences. It was demonstrated that simple linear variations of the control vectors can produce realistic articulatory movements and acoustic signals. In addition, it was shown that a simple command pattern can generate a complex velocity profile, due to biomechanical properties of the tongue. These results favor the use, in speech control modeling, of the EP hypothesis, that offers an efficient framework to understand the link between the phonological and physical levels of speech production. Further works will be focused on the modeling of muscle synergies, and on the study of the variability of motor control strategies related to prosodic changes.

Acknowledgements

The authors wish to thank Professor Lebeau from the Centre Hospitalier Universitaire de Grenoble for his assistance in their understanding of the anatomy of the tongue, as well as Pierre Badin and Gérard Bailly for their assistance with various aspects of the study. We also express our appreciation to Drs. David Ostry and Joe Perkell for their comments on

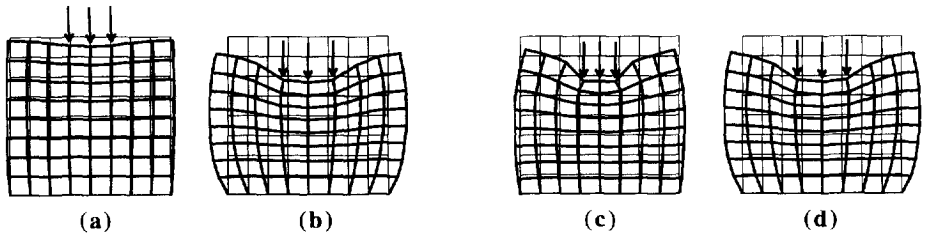

Fig. 13. The effects of the elastic constants (Young modulus on the left panels, Poisson ratio on the right ones). On the left, the structure shows larger deformations (b) in the case of a lower Young modulus. On the right side, the quasi-incompressibility is modeled ($\nu = 0.49$) (d), but is absent if the Poisson ratio is lower ($\nu = 0.2$) (c).

an earlier version of this manuscript, and to the reviewers, Drs. Kiyoshi Honda, Michael McClean, and Shinji Maeda, for their helpful comments and suggestions. We are very grateful to Linda Sherwood, who tried to make our English somewhat less difficult to read.

Appendix A. The Finite Element Method

The Finite Element Method (FEM) is the standard engineering technique for simulating the dynamical behavior of an object. It is a numerical procedure used for relating the forces acting on and within an object, to its subsequent motion and deformation (Schwartz, 1984; Zienkiewicz and Taylor, 1989). *Interpolation shape functions* are developed in order to distribute, across regions of interest, continuous material properties, such as mass, stiffness and deformation capabilities. Those functions relate the movement of one node of the body to the relative movements of all the other nodes. As rest configurations of the objects are often complex and not regular, a local coordinate system is used. Analytical transformations are then introduced, to transform this local coordinate system into the cartesian one. In the present version of the model, the body is assumed to be *isoparametric*, which means that element movements are interpolated with the same functions as in the transformation between coordinate systems.

The aim of this appendix is not a comprehensive description of the FEM (see Schwartz, 1984; Zienkiewicz and Taylor, 1989, for details). We only propose to show how this technique accounts for the effects of two elastic constants which characterize the physical behavior of the biomechanical structure, namely the *Young modulus E* and the *Poisson ratio* ν.

The Young modulus measures the stiffness of the material, that is to say the way the object responds to an external pressure force. Its value is low when the structure is easily malleable, and high in the case of a quite rigid body. On the other side, the Poisson ratio measures the way a deformation, in a given direction, can induce deformations in other directions. This elastic parameter is indirectly linked to the compressibility of the structure. A value close to zero means that the object accepts any deformation, in any direction, without any reaction in orthogonal directions. On the other hand, a value close to 0.5, describes, at the second order level, a quasi-incompressible structure, as an ε_x deformation along the x dimension, will induce $\varepsilon_y/2$ and $\varepsilon_z/2$ deformations, along the two orthogonal dimensions y and z (in the 3D case).

Fig. 13 shows the influence of those two elastic constants. A 2D square FE structure is put on a horizontal plane (its low nodes can thus move along the horizontal axis only). A constant vertical pressure force is applied at the top of the structure. The left panel of the figure shows the influence of the Young modulus: lower E value (right side of the panel) induces larger deformations. The right panel of the figure shows the effects of a change in the Poisson ratio value. Whereas the quasi-incompressibility modeling (ν close to 0.5) induces lateral deformations (right side of the panel), a lower Poisson ratio value has smaller effects (left side of the panel).

References

Adams, S.G., Weismer, G., Kent, R.D., 1993. Speaking rate and speech movement velocity profiles. Journal of Speech and Hearing Research 36, 41–54.

Adatia, A.K., Gehring, E.N., 1971. Proprioceptive innervation of the tongue. Journal of Anatomy 110, 215–220.

Badin, P., Fant, G., 1984. Notes on vocal tract computation. STL QPSR 2-3, 53–108.

Badin, P., Perrier, P., Boë, L.J., Abry, C., 1990. Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. Journal of the Acoustical Society of America 87, 1290–1300.

Badin, P., Gabioud, B., Beautemps, D., Lallouache, T.M., Bailly, G., Maeda, S., Zerling, J.P., Brock, G., 1995. Cineradiography of VCV sequences: articulatory-acoustic data for a speech production model. In: Proceedings of the 15th International Congress of Acoustics, Vol. IV, Trondheim, Norway, pp. 349–352.

Baer, T., Alfonso, P.J., Honda, K., 1988. Electromyography of the tongue muscles during vowels in /əpVp/ environment. Annual Bulletin of Research Institute of Logopedics and Phoniatrics 22, 7–19.

Beautemps, D., Badin, P., Laboissière, R., 1995. Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. Speech Communication 16 (1), 27–47.

Bothorel, A., 1975. Positions et mouvements de l'os hyoïde dans la chaîne parlée. Travaux de l'Institut de Phonétique de Strasbourg 7, 80–132.

Boyce, S.E., Krakow, R.A., Bell-Berti, F., Gelfer, C.E., 1990. Converging sources of evidence for dissecting articulatory movements into core gestures. Journal of Phonetics 18, 173–188.

Bunton, K., Weismer, G., 1994. Evaluation of a reiterant force-impulse task in the tongue. Journal of Speech and Hearing Research 37, 1020–1031.

Carré, R., Mrayati, M., 1991. Vowel-vowel trajectories and region modeling. Journal of Phonetics 19 (3/4), 433–444.

Cooper, S., 1953. Muscles spindles in the intrinsic muscles of the human tongue. Journal of Physiology 122, 193–202.

Duck, F.A., 1990. Physical Properties of Tissues: A Comprehensive Reference Book. Academic Press, London.

Dworkin, J.P., Aronson, A.E., Mulder, D.W., 1980. Tongue force in normals and in dysarthric patients with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research 23, 828–837.

Fant, G., 1960. Acoustic Theory of Speech Production. Mouton, The Hague.

Fant, G., 1992. Vocal tract area functions of Swedish vowels and a new three-parameter model. In: Ohala, J.J., Neaty, T., Derwing, B., Hodge, M., Wiebe, G. (Eds.), Proceedings of the International Conference on Spoken Language Processing, Edmonton, The University of Alberta, Vol. 1, pp. 807–810.

Feldman, A.G., 1966. Functional tuning of the nervous system with control of movement or maintenance of a steady posture – II Controllable parameters of the muscles. Biophysics 11, 565–578.

Feldman, A.G., 1981. The composition of central programs subserving horizontal eyes movements in man. Biological Cybernetics 42, 107–116.

Feldman, A.G., 1986. Once more on the Equilibrium-Point hypothesis (λ model) for motor control. Journal of Motor Behavior 18 (1), 17–54.

Feldman, A.G., Orlovsky, G.N., 1972. The influence of different descending systems on the tonic stretch reflex in the cat. Experimental Neurology 37, 481–494.

Feldman, A.G., Adamovich, S.V., Ostry, D.J., Flanagan, J.R., 1990. The origin of electromyograms – Explanations based on the Equilibrium Point Hypothesis. In: Winters, J., Woo, S. (Eds.), Multiple Muscles Systems: Biomechanics and Movement Organization. Springer, New York, pp. 195–213.

Flanagan, J.R., Ostry, D.J., Feldman, A.G., 1990. Control of human jaw and multijoints arm movements. In: Hammond, G. (Ed.), Cerebral Control of Speech and Limb Movements. Springer, New York, pp. 29–58.

Gambarelli, J., Guerinel, G., Chevrot, L., Mattei, M., 1977. Coupes Sériées du Corps Humain: Anatomie, Radiologie, Scanner. Masson, Paris.

Gerhardt, P., Frommhold, W., 1988. Atlas de Corrélations Anatomiques en Tomodensitométrie et Imagerie par Résonance Magnétique. Flammarion Médecine-Sciences, Paris.

Guenther, F.H., 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review 102 (3), 594–621.

Hashimoto, K., Suga, S., 1986. Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue. Journal of Acoustic Society of Japon (E) 7 (1), 39–46.

Hogan, N., 1984. An organizing principle for a class of voluntary movements. Journal of Neuroscience 4 (11), 2745–2754.

Honda, K., 1996. Organization of tongue articulation for vowels. Journal of Phonetics 24, 39–52.

Huxley, A.F., 1957. Muscle structure and theories of contraction. Prog. Biophys. Chem. 7, 255–318.

Kakita, Y., Fujimura, O., 1977. Computational model of the tongue: A revised version. Journal of the Acoustical Society of America 62, S15(A).

Kakita, Y., Fujimura, O., Honda, K., 1985. Computation of mapping from muscular contraction patterns to formant patterns in vowel space. In: Fromkin, V.A. (Ed.), Phonetic Linguistics. Academic Press, Orlando, FL, pp. 133–144.

Kiritani, S., Miyawaki, K., 1975. Computational model of the tongue. Journal of the Acoustical Society of America 57, S3(A).

Kiritani, S., Miyawaki, K., Fujimura, O., 1976. A computational model of the tongue. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics 10, 243–252.

Laboissière, R., Ostry, D.J., Feldman, A.G., 1996. The control of multi-muscle systems: Human jaw and hyoid movements. Biological Cybernetics 74 (3), 373–384.

Lallouache, M.T., 1990. Un poste "visage-parole". Acquisition et traitement de contours labiaux. Actes des 18èmes Journées d'Études sur la Parole, Montréal, pp. 27–28.

Lœvenbruck, L., 1996. Pistes pour le contrôle d'un robot parlant capable de réduction vocalique. Unpublished Doctoral Dissertation, Institut National Polytechnique de Grenoble, France.

MacNeilage, P., 1970. Motor control of serial ordering of speech. Psychological Review 77, 182–196.

Maeda, S., 1979. Un modèle articulatoire de la langue avec composantes linéaires. Actes des 10èmes Journées d'Étude sur la Parole, pp. 152–164.

Maeda, S., Honda, K., 1994. From EMG to formant patterns of vowels: The implication of vowel systems and spaces. Phonetica 51, 17–29.

McClean, M.D., Clay, J.L., 1995. Activation of lip motor units with variations in speech rate and phonetic structure. Journal of Speech and Hearing Research 38, 772–782.

Min, Y., Titze, I., Alipour, F., 1994. Stress-strain response of the human vocal ligament. NCVS Status and Progress Report 7, 131–137.

Miyawaki, K., 1974. A study on the musculature of the human tongue. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics 8, 23–50.

Morasso, P., 1981. Spatial control of arm movements. Experimental Brain Research 42, 223–227.

Munhall, K.G., Ostry, D.J., Parush, A., 1985. Characteristics of velocity profiles of speech movements. Journal of Experimental Psychology: Human Perception and Performance 11, 457–474.

Nelson, W.L., 1983. Physical principles for economies of skilled movements. Biological Cybernetics 46, 135–147.

Nittrouer, S., Munhall, K., Kelso, J.A.S., Tuller, B., Harris, K.S., 1988. Patterns of interarticulator phasing and their relation to linguistic structure. Journal of the Acoustical Society of America 84, 1653–1661.

Oka, S., 1974. Rheology-Biorheology. Shokabo, Tokyo, pp. 454–456.

Ostry, D.J., Munhall, K.G., 1985. Control of rate and duration of speech movements. Journal of the Acoustical Society of America 77, 640–648.

Ostry, D.J., Flanagan, J.R., 1989. Human jaw movement in mastication and speech. Archs Oral Biol. 34 (9), 685–693.

Ostry, D.J., Flanagan, J.R., Feldman, A.G., Munhall, K.G., 1992. Human jaw movement kinematics and control. In: Stelmach, G.E., Requin, J. (Eds.), Tutorials in Motor Behavior II. Elsevier, Amsterdam, pp. 646–660.

Payan, Y., Perrier, P., Laboissière, R., 1995. Simulation of tongue shape variations in the sagittal plane based on a control by the Equilibrium-Point hypothesis. In: Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Sweden, Vol. 2, 474–477.

Pelorson, X., Vescovi, C., Castelli, E., Hirschberg, A., 1995. Aeroacoustical aspects of vocal tract excitation: Contribution to articulatory synthesis. In: Proceedings of the 15th International Congress of Acoustics, Trondheim, Norway, pp. 505–508. .

Perkell, J.S., 1974. A physiologically-oriented model of tongue activity in speech production. Unpublished Doctoral Dissertation Thesis, Massachusetts Institute of Technology, Boston, MA.

Perkell, J.S., 1996. Properties of the tongue help to define vowels categories: Hypotheses based on physiologically-oriented modeling. Journal of Phonetics 24, 3–22.

Perrier, P., Lœvenbruck, H., Payan, Y., 1996a. Control of tongue movements in speech: The Equilibrium Point hypothesis perspective. Journal of Phonetics 24, 53–75.

Perrier, P., Ostry, D.J., Laboissière, R., 1996b. The Equilibrium Point Hypothesis and its application to speech motor control. Journal of Speech and Hearing Research 39 (2), 365–377.

Schwartz, H.R., 1984. Finite Element Methods. Academic Press, London.

Smith, K.K., Kier, W.M., 1989. Trunks, tongues, and tentacles: Moving with skeletons of muscle. American Scientist 77, 29–35.

Soechting, J.F., Lacquaniti, F., 1981. Invariant characteristics of pointing movement in man. Journal of Neuroscience 1, 710–720.

Sock, R., Löfqvist, A., 1995. Some timing constraints in the production of bilabial stops. Journal of Phonetics 23, 129–138.

Walker, L.B., Rajagopal, M.D., 1959. Neuromuscular spindles in the human tongue. Anatomical Records 133, 438.

Weddell, G., Harpman, J.A., Lambley, D.G., Young, L., 1940. The innervation of the musculature of the tongue. Journal of Anatomy 74, 255–267.

Wells, J.B., 1965. Comparison of mechanical properties between slow and fast mammalian muscles. Journal of Physiology 178, 252–269.

Wilhelms-Tricarico, R., 1995. Physiological modeling of speech production: Methods for modeling soft-tissues articulators. Journal of the Acoustical Society of America 97 (5), 3085–3098.

Wilhelms-Tricarico, R., 1996. A biomechanical and physiologically-based vocal tract model and its control. Journal of Phonetics 24, 23–38.

Zienkiewicz, O.C., Taylor, R.L., 1989. The Finite Element Method. Basic Formulation and Linear Problems. MacGraw-Hill, Maidenhead, UK.