



**HAL**  
open science

## Biomechanical models to study speech

Pascal Perrier, Yohan Payan, Stéphanie Isabelle Buchaillard, Mohammad Ali Nazari, Matthieu Chabanas

► **To cite this version:**

Pascal Perrier, Yohan Payan, Stéphanie Isabelle Buchaillard, Mohammad Ali Nazari, Matthieu Chabanas. Biomechanical models to study speech. *Faits de langues*, 2011, 37, pp.155-171. hal-00619698

**HAL Id: hal-00619698**

**<https://hal.science/hal-00619698v1>**

Submitted on 6 Sep 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Biomechanical models to study speech

Pascal Perrier, Yohan Payan, Stéphanie Buchaillard,  
Mohammad Ali Nazari & Matthieu Chabanas\*

## 1. INTRODUCTION

Linguistic oral communication is a cognitive task in which humans share information by producing oro-facial gestures and generating acoustic signals. These gestures and signals are perceived, classified and interpreted in relation to a code, the linguistic code made of units of various sizes and of rules that specify how units can be assembled. In line with many authors, we think that linguistic units and rules have not been elaborated independently of the physical and cognitive systems that vehicle them in oral communication among humans (Liljencrants & Lindblom, 1972; Lindblom, 1983, 1984, 1986; Ohala, 1983; Browman & Goldstein, 1986, 2000; Boë et al., 2000; Schwartz et al, 1997; Schwartz et al, 2002; Schwartz et al, 2007; MacNeilage, 1998). We share their opinion in that the structures of languages and their diachronic evolutions are strongly influenced by the characteristics of the speech production and speech perception systems (Fuchs et al., 2007, Fuchs & Perrier, 2009). This point of view is in opposition with some major linguistic theories (see among others Chomsky & Halle, 1968, or Prince & Smolensky, 2004)

More specifically, for speech production, we think that anatomical and mechanical properties of the speech apparatus have to be taken into account in the study of languages. They should be considered in conjunction with general motor control mechanisms and strategies underlying the generation of skilled gestures in humans. The physical patterns carrying linguistic information in speech are typically associated with time varying articulatory and spectral patterns. These patterns are generated by displacements of the jaw, the tongue and the velum, by shaping of the deformable tongue and lips, and/or by vibrations of the vocal folds. These physical phenomena are deeply constrained, limited and determined by the physical nature of these articulators. All these constraints are likely to influence the structure of the phonological systems. Thus, Hamann (2002) has suggested that an explanation for the absence of non-velarized retroflexes in the inventory of world's languages could be the

---

\* Pascal Perrier : DPC/GIPSA-lab, CNRS, Grenoble-INP, UJF, Université Stendhal.  
Courriel : Pascal.Perrier@gipsa-lab.grenoble-inp.fr  
Yohan Payan : TIMC-IMAG, CNRS, UJF, Grenoble-INP. Courriel :  
Yohan.Payan@imag.fr  
Stéphanie Buchaillard : DPC/GIPSA-lab, CNRS, Grenoble-INP, UJF, Université  
Stendhal. Courriel : stephanie.buchaillard@gmail.com  
Mohammad Ali Nazari : DPC/GIPSA-lab, CNRS, Grenoble-INP, UJF, Université  
Stendhal. Courriel : Mohammad.Nazari@gipsa-lab.grenoble-inp.fr  
Matthieu Chabanas : DPC/GIPSA-lab, CNRS, Grenoble-INP, UJF, Université Stendhal.  
Courriel : Matthieu.Chabanas@gipsa-lab.grenoble-inp.fr

articulatory impossibility to generate retroflexion without tongue retraction. In the same vein, Hamann & Fuchs (2010) have suggested that the emergence of a voiced retroflex stop from a voiced alveolar or dental stop could arise from the specific articulatory requirements for voiced alveolar or dental stops.

Similarities between speech rhythms (Ohala, 1975) and natural oscillation frequencies of the mandible let think that timing in speech could be constrained by inertial and modal properties of the articulators (Morimoto et al., 1984). Articulatory paths and formant transitions between sounds are determined by the deformation capabilities of the tongue and of the lips (Perrier et al., 2003).

Hence, a careful analysis of the physical properties of the speech apparatus, including an anatomical, biomechanical and aerodynamic characterization should contribute significantly to the understanding of the structures of languages and of their evolutions. This has been well understood a long time ago by researchers like Perkell or Fujimura, who initiated pioneer works in physiological modeling of the tongue based on a description of muscle anatomy, muscle force generation mechanisms and tongue tissues mechanics (Perkell, 1974; Kiritani et al., 1976; Kakita & Fujimura, 1977). Since these pioneer works, the tremendous increase of computational capabilities in computers and the improvements of anatomical, physiological and neuro-physiological knowledge of the oro-facial sphere have permitted the elaboration of more sophisticated and more realistic models (for the tongue: Kakita *et al.*, 1985; Hashimoto & Suga, 1986; Wilhelms-Tricarico, 1995; Payan & Perrier, 1997; Sanguineti *et al.*, 1998; Wilhelms-Tricarico, 2000; Dang & Honda, 2004; Gerard et al., 2006; Fujita et al., 2007; Buchaillard et al., 2009; for the face: Terzopoulos & Waters, 1993; Lucero & Munhall, 1999; Gomi et al., 2006; Kim & Gomi, 2007; Sifakis et al., 2006; Nazari et al., 2010; for the jaw: Flanagan et al., 1990, Laboissière et al., 1996, Stavness et al., 2010).

In this paper, the methodology elaborated in our group for the design of biomechanical models of speech articulators (more specifically tongue and lips) in the aim to study speech production is explained. The basic principles that guided our work are close to those proposed by Wilhelms-Tricarico (1995) or Sifakis et al. (2006). Our models are continuous ones based on finite element structures, the meshes, as opposed to discrete models based on spring-mass networks such as those of by Kim & Gomi (2007), Dang & Honda's (2004), Terzopoulos & Waters's (1993). Muscle activation affects the mechanical properties of the mesh elements associated with the activated muscle; the mechanical properties of the mesh elements are non-linear. As compared to Wilhelms-Tricarico's or Sifakis et al.'s models, our muscle models are functional and then less accurate. However, in our models the interaction of the soft tissues with the external structures, the teeth, the palate, the cranial bones are modeled with contacts, and this was not the case in these two reference models.

In the current paper, the methodology that guided the design of oro-facial biomechanical models in Gipsa-lab is explained and examples of the most recent models are presented. Then, results are described that were obtained with simulations and are related to important issues in phonetics.

## 2. METHODOLOGY FOR THE DESIGN OF BIOMECHANICAL MODELS OF SPEECH ARTICULATORS

From a methodological point of view, our approach is constrained by two important and somewhat contradictory requirements: (1) since the aim of our work is to use models in order to infer and to understand the impact of physical phenomena underlying speech production onto phonetic patterns of languages, these models have to be as realistic as possible; (2) due to limitations in terms of computational costs and computational accuracy, numerical models of physics have to be simplified and inaccurate descriptions of reality have to be tolerated as long as they are well identified and explained. Hence, our methodology is the result of a compromise that consists in representing major anatomical, physiological, neuro-physiological and physical features and in suggesting or tolerating a number of simplifications in the description of some other properties that are assumed to have impacts of secondary importance.

### 2.1 3D geometry

The 3D geometry of the modeled articulators has been extracted from medical images collected from human subjects thanks to Magnetic Resonance Imaging (MRI) techniques, cineradiographies and CT-scans (Figure 1). Both in MRI and in CT scans, subjects are in supine position, which influences tongue geometry due to the orientation of the gravity field. Cineradiographies are collected with the subject's head in vertical position, and they are used to correct the geometry extracted from 3D scans and to remove the effect of gravity in supine position, by taking into account tongue contours in the mid-sagittal plane (Buchaillard, 2007)

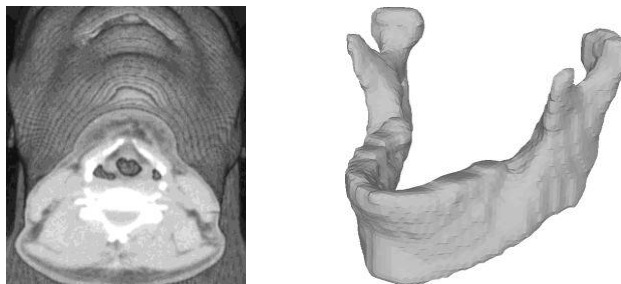


Figure 1: Example of a 3D CT Scan (left) and of a mandible geometry (right) extracted from CT scan data. In the CT-scan, white regions correspond to bony structures and dark gray regions to empty regions (air). (Buchaillard, 2007)

### 2.2. Finite Element Modeling

Modeling of soft tissues' mechanical behaviors, such as those of the tongue, the lips, the face or the velum, requires the application of physical laws specified

by continuum mechanics. These laws account, among others, for relations between stress and strain, and between forces and displacements, as well as for balance conditions of the soft body under the influence of external actions. An efficient numerical framework to implement these laws is the Finite Element Method (FEM) proposed by Turner *et al.* (1956). FEM requires sub-dividing the body into a set of small elements (the FEM mesh) determined by nodes and edges. Mechanical equations are solved for each element individually and rules are applied to assemble elements back into the whole body in order to determine its global deformation and displacement. Other techniques exist to account for soft tissues mechanics, which rely on a discrete representation of the body based on a series of connected springs and masses. These techniques have been shown to be computationally less demanding and therefore faster than FEM, but they also have been shown to be less accurate and less stable (Piterman, 2004).

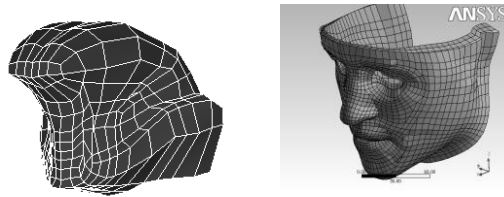


Figure 2: FEM meshes used for the tongue model (left, Buchaillard *et al.*, 2009) and the face model (right, Nazari *et al.*, 2010)

In our model we have used a well-acknowledged commercial FEM Package called ANSYS<sup>TM</sup>. It guarantees the best possible accuracy of numerical modeling. FEM meshes defined for the tongue and the face models are shown in figure 2. They are made of hexahedral elements. The orientations and sizes of the elements have been determined in order to facilitate the modeling of muscle anatomy within each articulator and in the face (see below).

### 2.3. Anatomy and physiology

Face and tongue anatomies are very complex. Mucosa, skin tissues, fat tissues, muscular tissues, ligaments and collagens are mixed with each other. Their respective boundaries are not clear and all these different kinds of tissues are unevenly distributed spatially. In many parts of the tongue fibers belonging to different muscles and oriented along various directions are interwoven. Our models do not account in details for this complexity. Based on dissections results (Miyawaki, 1974; Takemoto, 2001), on MRI images (Gilbert & Napadow, 2005) on data provided by the Visual Human Project (National Library of Medicine, USA, [www.nlm.nih.gov/research/visible](http://www.nlm.nih.gov/research/visible)), and on anatomical atlases (Netter, 2004), we have proposed a simplified description of the anatomy, in which muscles are represented by a set of contiguous elements within the FEM mesh and by a set of macrofibers oriented for each muscle along the main direction of

its fibers. The mesh models the elastic behavior of the tissues via a constitutive law representing the relation between stress and strain. Both linear and nonlinear laws have been used, which is done in the linear case simply with the specification of the Young's modulus and the assumption of incompressibility. The macrofibers are connected to the Finite Element mesh and they exert strain on it. The shape of the FEM mesh elements has been adapted, in order to fit as well as possible the global geometry of each muscle (main direction and cross-section). The connections of the tongue with the mandible and the hyoid bone and of the face tissues with the skull have been carefully described. The insertions of the muscles on the bony structures and within the soft tissues have been also determined with accuracy. They are of particular importance, since they determine the main directions of the macrofibers and, consequently, of the muscle forces exerted on the FEM mesh.

In its last version the 3D model of the tongue includes eleven independent muscle units, controlling the intrinsic muscles (the posterior, medium, and anterior parts of the Genioglossus, the Superior Longitudinalis, the Inferior Longitudinalis, the Verticalis, and the Transversalis), the extrinsic muscles (Hyoglossus and Styloglossus) and the mouth floor muscles (Geniohyoid and Mylohyoid). The face model includes also eleven independent muscle units, the Levator Labii Superioris Alaeque Nasi, Depressor Labii Inferioris, the Levator and Depressor Anguli Oris, the Zygomaticus, the Risorius, the Buccinator, the Mentalis and the Peripheral and Marginal parts of the Orbicularis Oris. Examples of these muscle implementations are given in figure 3.

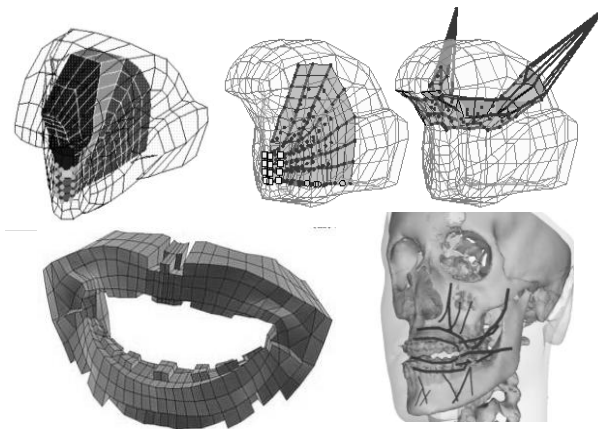


Figure 3: Examples of anatomical muscle implementations. Top panel: left, localization of the anterior (dark gray), medium (light gray) and posterior (gray) parts of the genioglossus in the tongue; right, macrofibers for the posterior genioglossus and the styloglossus in the tongue model (Buchallard et al., 2009). Bottom panel: left, volumetric description of the orbicularis oris in the face model; right: macrofibers for the main muscles in the face model (left side only) (Nazari et al., 2010)

#### 2.4. Mechanical aspects

Mechanical aspects include mass, damping factors, elastic properties of tissues, muscle force generation mechanisms, as well as boundary conditions (i.e., in our models, contacts between structures). The mass depends on material's mass density and volume. For the bony parts, the density is  $2000\text{kg/m}^3$ . For soft tissues, the density was chosen to be close to the density of water, i.e.  $1040\text{kg/m}^3$ . Damping factors have been chosen to be close and greater than critical damping in the range of movement frequencies observed in speech (from 1 to 20Hz). This allows the articulators moving rapidly, without generating physical instabilities.

Elastic properties of soft tissues were determined from measures made on tongue's and cheek pieces removed from a fresh cadaver (Gerard et al., 2005), and from complementary measures made on living subjects (Schiavone et al., 2008). These measures provide information about the relation between the intensity of an external stress applied to the tissues and the corresponding tissues deformation. This relation was found to be non-linear: a given increase of the stress has a smaller impact when deformation is large than when it is small (i.e. the stiffness increases for large deformation). This could be modeled by elements made of hyperelastic material (Gerard et al., 2005). In our models, these mechanical properties are modeled isotropically: they are exactly the same in all directions. This isotropic approach was chosen for sake of simplicity. It should be acknowledged that this does not match experimental results. Indeed, muscles have been shown to be rather transversely isotropic: their mechanical properties are fairly identical in directions orthogonal to the long muscle fibers, but they are different in the direction of these fibers.

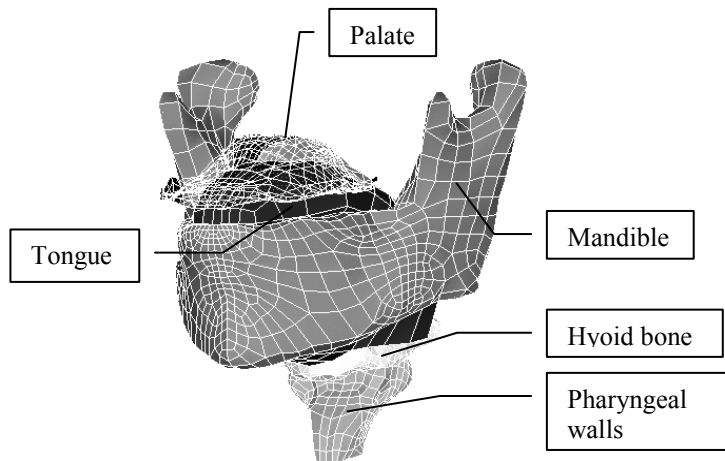


Figure 4: The 3D vocal tract model (Buchaillard et al., 2009)

The tongue model is embedded in a vocal tract and connected to the jaw and the hyoid bone. Vocal tract boundaries are the mandible, the hard and the soft palate, and the pharyngeal walls (Figure 4). The geometry of the mandible, the teeth, the hard palate, the soft palate and the pharyngeal walls were extracted from MRI images (Badin et al., 2002). In the current version of the model the mandible is fixed. In addition the soft palate and the pharyngeal walls are represented as solid structures. The hyoid bone is mobile and connected with springs to lower bony structures in the laryngeal and chest regions (Gerard et al., 2006, Buchaillard et al., 2009). Boundary conditions correspond to contacts of the tongue blade and tongue tip with palate and upper teeth, and of the lower part of the tongue tip with the lower teeth and inner surface of the mandible (Buchaillard et al., 2009). The face model includes contacts between upper and lower lips, as well as contact between lips and teeth (Nazari et al., 2010).

### 2.5. Muscle model

A muscle generates a force along the direction of its long fibers. This force tends to shorten the muscle and varies non-linearly with the length and the shortening rate of the long fibers (McMahon, 1984). In addition, when a muscle is activated, it becomes stiffer in the direction orthogonal to its long fibers (McMahon, 1984) (the so-called “*stress stiffening effect*”). In our modeling approach, these phenomena are accounted for (Nazari et al., 2010). A muscle force is generated according to the Equilibrium-Point Hypothesis (Feldman, 1986), which specifies that force varies as an exponential function of the muscle length (see below). In addition, this force decreases when the shortening rate increases. The stress stiffening effect is modeled thanks to changes in the relation between force and deformation.

### 2.6. Motor control

The choice for a motor control model was strongly influenced by our hypotheses about the relations between phonological units and their physical realizations. We do think that the discrete nature of the phonological inputs is mirrored in the motor control variables, and that the physical continuum of the articulatory movements and of the acoustic signal is the result of the interaction between discrete sequences of motor commands and the physical properties of the speech apparatus (Perrier et al., 1996a). Thus, our model of speech motor control is a “target-based” model, in which the targets are specific sets of motor commands associated with the phonological units. To implement this view of speech motor control, we used Feldman’s Equilibrium-Point Hypothesis and its associated  $\lambda$ -model (Feldman, 1986). According to this hypothesis a movement is generated from a discrete sequence of target motor commands from which the continuous time variation of the motor control variables is determined by shifting them at a constant rate between successive target values. The motor control variables are called  $\lambda$  and they are muscle threshold lengths: when a muscle is shorter than the  $\lambda$  value, it does not generate any active force (it behaves like a



passive spring); when a muscle is longer than  $\lambda$ , it generates a force that tends to shorten it and to bring it back to length  $\lambda$ . Muscle force increases exponentially with the difference between the actual muscle length and the threshold  $\lambda$ . When two or more muscles are active, the resulting position of the articulator corresponds to a mechanical equilibrium between the different muscles forces. Feldman's hypothesis suggests that the mechanical equilibrium is controlled by the Central Nervous System via the specification of the  $\lambda$  thresholds of all active muscles.

Thus physical targets in speech production are not specified as articulatory positions, but as dynamical equilibrium positions. This property is very important and has provided a powerful modeling framework to account for the physical variability observed for a given phonological input according to speaking style, speaking rate or phonetic context (Perrier et al., 1996a; Perrier et al., 1996b) (see also below).

### 2.7. Evaluation

The evaluation of the models is based on comparisons of simulations results with experimental data collected on human speakers. Comparisons do not consist of a computation of any kind of average error between experimental trajectories and simulated ones. Indeed, in spite of the fact that our biomechanical models are based on data collected on human speakers and account as carefully as possible for physiological and physical properties, such a comparison would not make much sense. Some physical parameters in the model cannot be actually measured on living subjects, as for example the damping factors, or the variation of stiffness with muscle activation, or the global level of muscle activation. In addition the model is symmetrical in reference to the mid-sagittal plane, which is never really the case in humans. In addition, in spite our wish to be as accurate and realistic as possible, we had to tolerate some simplifications, such as the hypothesis of tissue isotropy or the application of muscle forces via macrofibers. These discrepancies with real human beings are not supposed to induce strong differences between simulations and data. However, they do not allow a point-to-point comparison of trajectories. Rather, comparisons are based on a number of more macroscopic factors that are key factors in terms of motor control or in terms of speech production, such as vocal tract shape, movement amplitude, movement duration, amplitude of the velocity peak within a movement, velocity profiles (time variation of the velocity during a movement), maximum level of force, position of acceleration peaks....

This will be illustrated in the section below.

## 3. EXAMPLES OF SIMULATION RESULTS RELATED TO ISSUES IN PHONETICS

In this section, some examples of results will be presented, that have potential implications for the studies of languages.

### 3.1. Tongue muscle activation for vowel /i/ - Possible consequences of control inaccuracies

With the vocal tract model presented Figure 4, 300 simulations have been carried out, in order to find out the muscle activations underlying the production of the main French oral vowels. These simulations have been assessed in the articulatory domain by comparing real 3D tongue shapes with the simulated ones, as well as in the acoustic domain by looking at the frequencies of the three first spectral maxima (formants) of the acoustic signal generated from the simulated tongue shape. The three formants are extracted from the synthetic signal generated from the 3D vocal tract shape with a Kelly-Lochbaum model implemented by Story (Story, 2005). Initial activation patterns have been proposed on the basis of electromyographic data collected by Miyawaki *et al.* (1975) and Baer *et al.* (1988) for Japanese and English vowels. From these initial patterns, variations were provided to the motor commands in order to generate realistic tongue shapes. To do so, our approach was, on the one hand, to make the simulated tongue shapes in the midsagittal plane as close as possible to tongue shapes classically presented for French oral vowels in the literature and observed by means of cineradiographic data (see for example Bothorel *et al.*, 1986). On the other hand, we tried to get a good match between the synthetic formant values generated from the simulated tongue shape and the formant patterns classically given in the literature.

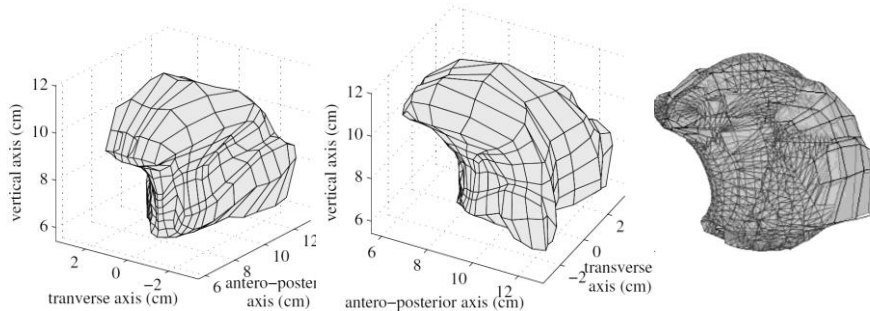


Figure 5: Tongue shape simulated for vowel /i/. Left panel: view from the front; Middle panel: View from the back; Right panel superimposition with the MRI data recorded from a native speaker of French (triangular mesh) (Buchaillard *et al.*, 2009)

The tongue shape thus obtained for vowel /i/ is depicted on Figure 5 (Buchaillard *et al.*, 2009). The corresponding formant values are  $F1=321\text{Hz}$ ,  $F2=2095\text{Hz}$  and  $F3=2988\text{Hz}$ . They correspond very well to the average values proposed for this vowel in French in Calliope (1984). The classical articulatory characteristics of this vowel are well replicated: the highest point of the tongue is located in its front part (alveo-palatal region), the back part is clearly pushed forward and it widens the pharyngeal cavity, the front part is grooved along the mid-sagittal direction. This is in agreement with general descriptions of the 3D

shape of vowel /i/ (for example in Lundberg & Stone, 1999) and with 3D data measured for this vowel with anatomic MRI on the native speaker of French<sup>1</sup> who served as reference to build up the tongue model.

The main tongue muscles underlying the production of this tongue shape are the posterior part of the genioglossus (GGp), its anterior part (GGa), the styloglossus (Sty) and the transversalis (Trans). In addition contacts are observed between the hard palate and the upper dental arch on one hand and the tongue lateral borders on the other hand. Our simulations reveal that GGp and Sty activations are responsible for tongue elevation, that GGa activation also allows the anterior positioning, and that the grooving of the anterior part is due to GGa activation. Transversalis activation facilitates these different shapings by limiting tongue deformation in the transversal direction. These findings are consistent with the EMG studies by Baer et al. (1998). As compared to these important and basic experimental results, our model brings a significant amount of new information, since it enables to understand the contribution of each muscle in the tongue shape of this vowel

An analysis of the consequences of small variations of muscle activations has shown that they essentially affect the size of the constriction due to variations of the grooving amplitude, the front positioning of the tongue being stabilized by the contacts with the palate and the teeth (Buchaillard et al., 2009). This finding shows that the correct production of vowel /i/ requires an accurate coordination of GGa and Trans muscles, in order to produce the appropriate grooving amplitude. Any inaccuracy in this coordination, due either to changes in motor commands or to variation in the relative timing of these commands, is likely to induce a transformation from a vowel /i/ to a more constricted articulation. This kind of inaccuracy could be at the origin of linguistic changes, where the vowel /i/ becomes a glide /j/ in vowel-to-vowel transitions (see Ladefoged & Maddieson, 1996; Chitoran & Hualde, 2007; Padgett, 2008)

### 3.2. *Role of inertia and dynamics: vowel reduction*

Vowel reduction associated with an increase in speaking rate is a well-known variation particularly well illustrated by the experimental study of /dud/ variability carried out by Lindblom (1963). On the basis of simulations run with a mass-spring model, Lindblom (1963) could reproduce the measured spectrographic patterns by decreasing the time allowed for the vowel production between the two consonants. Doing so Lindblom suggested that articulatory inertial and dynamical properties would be the main factors of vowel reduction in fast speaking rate condition. However, the question remained whether realistic vowel reduction patterns can be obtained without changing the target articulatory commands underlying the production of the inter-consonantal vowel. Loevenbruck & Perrier (1993), using a slightly more realistic spring mass model controlled by the Equilibrium Point Hypothesis, suggested that vowel reduction can be obtained without any change in target vowel commands, just by reducing

---

<sup>1</sup> Courtesy of Pierre Badin, Gipsa-lab, CNRS, Grenoble, France.

movements' durations. A more accurate assessment is now possible by using our realistic biomechanical models.

For this study, we used the two dimensional version of the vocal tract model. This model has been extensively described in Payan & Perrier (1997) and Perrier et al. (2003). As compared to the description that is provided above of the more recent 3D model, some points should be made clear though. The Finite Element mesh is a representation of the projection of the tongue in the mid-sagittal plane. Just as in the 3D model the internal geometry of the mesh has been designed to facilitate the anatomical representation of the muscles. However, the restriction of the mesh to a 2D representation has two noticeable consequences. First, only the muscles that have a clear and direct impact on the tongue contour in the mid-sagittal plane are modeled. These muscles are the posterior and anterior parts of the Genioglossus, the Hyoglossus, the Styloglossus, the Verticalis and the inferior and superior Longitudinalis. The role of the transversalis is consequently not accounted for. Second, the non-compressibility constraint is not implemented in relation to the 3D volume of the tongue, but in relation to the 2D surface of the Finite Element mesh. In addition, the elastic properties of the mesh elements are not described by a constitutive law relating stress and strain, but directly via the Young Modulus. Hence, the tissue stiffening associated with muscle activation is modeled by a linear increase of the Young Modulus with muscle activation (according to Feldman's model the activation is the difference between the actual muscle length and the motor command  $\lambda$ , and takes also into account the rate of muscle length change).

The study of the vowel reduction phenomenon with this model consisted in observing the impact of speaking rate increase on the tongue shape reached for the vowel. It has been assessed for /kik/ and /kak/ sequences. The targets for /k/, /a/ and /i/ have been determined with an optimal planning model described in Perrier & Ma (2008). This model minimizes the articulatory effort while ensuring that targets in the acoustic domain are reached. It takes into account the fact that the jaw is fixed in the model. Results are presented in Figure 6. They are evaluated with respect to the hypothesis suggested by Loevenbruck & Perrier (1993) on the basis of the tongue shape actually reached for vowels /i/ and /a/. These tongue shapes are compared to the rest position of the tongue model (close to a schwa) and to the other tongue shapes generated with this model for the other French oral vowels. In addition, general information about "standard" vocal tract configurations measured on human speakers in the mid-sagittal plane (such as the X-ray views available in Bothorel et al., 1986) is taken into account. For both speaking rates the timings of motor commands are the same for both sequences. For each sequence target motor commands are identical in both speaking rates. It can be observed on Figure 6 that the impact of speaking rate increase is different according to the sequence. For fast speaking rate, the tongue shape reached for /a/ is very close to the central schwa configuration, without any clear constriction in the vocal tract (right top panel). Under the same conditions, the shape reached for /i/ is close to an /e/ configuration. Indeed, the right bottom panel of Figure 6 shows that this tongue shape still corresponds to a high and front vowel. However, as shown by the global orientation of the nodes

trajectories in the palatal region from /k/ to /i/, this tongue shape is less high and less front than under normal speaking rate conditions (Figure 6, bottom panel, left),

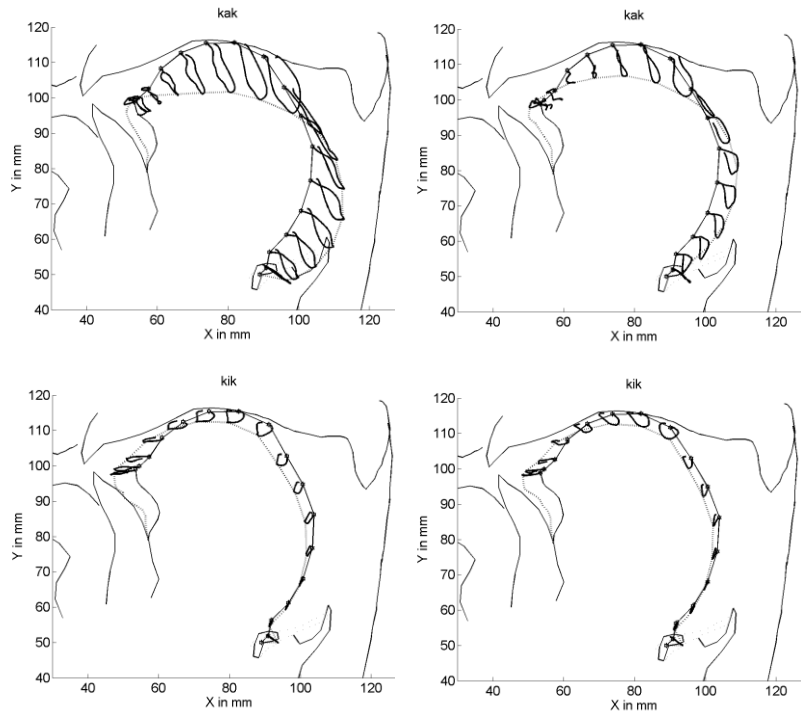


Figure 6: Impact of speaking rate on /kVk/ articulatory patterns simulated with a 2D biomechanical model of the vocal tract.

Top panel: sequence /kak/, bottom panel: sequence /kik/;

Left: normal speaking rate; right: fast speaking rate.

In each plot the mid-sagittal tongue contour is represented for the first consonant /k/ (upper contour) and for the vowel (lower contour). Trajectories of nodes located on the upper tongue contour are depicted for the whole /kVk/ sequence

These results are consistent with the acoustical observations made by Gendrot & Adda-Decker (2009) from a large corpus of multi-speaker spontaneous speech. Thus, the simulations made with a realistic biomechanical model confirm the trends of the observations made by using more simple models, and

make them more specific thanks to the integration of global constraints on tongue deformation.

The fact that vowel dependent articulatory (and acoustic) variability can be realistically accounted for with constant underlying motor targets, just by varying the timing of the motor command patterns is interesting and important in the context of the study of the interface between (invariant) phonology and (variable) phonetics.

#### 4. CONCLUSIONS

Elaborating realistic biomechanical models of the speech articulators is a long-term task, requiring the collection of many data of different kinds about the anatomical, physiological and mechanical characteristics of the peripheral speech production system, and necessitating the use of complex and powerful mathematical methods. At first glance it seems very far from major linguistic issues. However in the context of substance-based approaches of language studies, this methodology contributes efficiently to quantitatively evaluate possible interactions between the structures of languages, their variations and their evolutions and the physical characteristics of the peripheral speech production apparatus. The study of the muscle activations underlying the vowel /i/ and of the sensitivity of this vowel to variations in the activations' amplitudes and in their time coordination has provided a basis for a further study of sounds' changes that could originate in these variations. The study of vowel reduction has shown that variable articulatory and acoustic patterns can be found for constant motor commands simply by varying their timing. Thus, biomechanical models of speech production offer a powerful research context to quantitatively study the interface between invariant phonological inputs and variable phonetic outputs, with a possible emphasis of the role (and the status) of the prosody in this interface.

#### 5. ACKNOWLEDGMENTS

This work has been funded by grants provided by the CNRS (POPAART Project, PR2 French-Germany Program), by the French-German University (Project Pilios) and by the Agence Nationale de la Recherche (Project Skullspeech, ANR-08-BLAN-0272).

#### 6. REFERENCES

- Badin P., Bailly G., Réveret L., Baciú M., Segebarth C. & Savariaux C., 2002, Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images, *J. Phonetics* 30(3), p. 533-553.
- Baer T., Alfonso P. J. & Honda K., 1988, Electromyography of the tongue muscles during vowels in /əpVp/ environment, *Annual Bulletin of the Research Institute of Logopedics and Phoniatric* 22, Tokyo University, p. 7-19, URL <http://www.umin.ac.jp/memorial/rilp-tokyo/>.

- Boë L.-J., Vallée N., Badin P., Schwartz J.-L. & Abry C., 2000, Tendancies in phonological structures: the influence of substance on form, *Bulletin de la Communication Parlée* 5, Gipsa-lab, Grenoble-INP, Grenoble, p. 35-55
- Bothorel A., Simon P., Wioland F. & Zerling, J.-P., 1986, *Cinéradiographie des voyelles et des consonnes du français (Cineradiography of vowels and consonants in French)*, Institut de Phonétique, Université Marc Bloch, Strasbourg, France.
- Browman C. P. & Goldstein L. M., 1986, Towards an articulatory phonology, *Phonology Yearbook* 3, p. 219-252.
- Browman C. P. & Goldstein L. M., 2000, Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5, Gipsa-lab, Grenoble-INP, Grenoble, p. 25-34.
- Buchaillard S., 2007, *Activations musculaires et mouvements linguaux : modélisation en parole naturelle et pathologique*, Unpublished Ph.D. Thesis, Université Joseph Fourier, Grenoble, France (URL: [http://tel.archives-ouvertes.fr/docs/00/28/99/15/PDF/these\\_SBuchaillard.pdf](http://tel.archives-ouvertes.fr/docs/00/28/99/15/PDF/these_SBuchaillard.pdf))
- Buchaillard S, Perrier P & Payan Y. 2009, A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning, *J. Acoust. Soc. Am.* 126(4), p. 2033-2051
- CALLIOPE, 1989, *La Parole et son traitement automatique*. Paris : Masson
- Chitoran I. & Hualde J.I., 2007, From hiatus to diphthong: the evolution of vowel sequences in Romance, *Phonology* 24, p. 37-75
- Chomsky N. & Halle M., 1968, *The Sound Patterns of English*. New York: Harper & Row.
- Dang J. & Honda K., 2004, Construction and control of a physiological articulatory model, *J. Acoust. Soc. Am.* 115(2), p. 853-870.
- Feldman A.G., 1986, Once more on the Equilibrium-Point hypothesis ( $\lambda$  model) for motor control, *Journal of Motor Behavior* 18, p. 17-54.
- Flanagan J.R., Ostry D.J. & Feldman A.G., 1990, Control of human jaw and multijoints arm movements, in G. Hammond (ed.), *Cerebral control of speech and limb movements*, Springer Verlag, New-York, p. 29-58.
- Fuchs S., Pompino-Marschall B. & Perrier P., 2007, Is there a biological grounding of phonology? Determining factors, optimization and Communicative usage, *Proceedings of the XVI<sup>th</sup> International Congress of Phonetic Sciences*, Saarbrücken, Germany, p. 219-222.
- Fuchs S., & Perrier P., 2008, Understanding speech production: The PILIOS approach, *Revue française de linguistique appliquée*, XIII 2, p. 35-44
- Fujita S., Dang J., Noriko S. & Kiyoshi H., 2007, A computational tongue model and its clinical application, *Oral Science International*, p. 97-109,
- Gendrot C. & Adda-Decker M., 2009, Influence du contexte consonantique et de la durée des voyelles sur la centralisation des voyelles orales en français, in M. Embarki & C. Dodane, *La Coarticulation, des Indices aux Représentations*, Paris : l'Harmattan.
- Gerard J.-M., Ohayon J., Luboz V., Perrier P. & Payan Y., 2005, Non linear elastic properties of the lingual and facial tissues assessed by indentation technique. Application to the biomechanics of speech production, *Medical Engineering & Physics* 27, p. 884-892.
- Gerard J.-M., Perrier P. & Payan, Y., 2006, 3D biomechanical tongue modelling to study speech production, in J. Harrington & M. Tabain (eds.) *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press, New-York, p. 85-102.

- Gilbert R. J., Napadow V. J., Gaige T. A. & Wedeen V. J., 2007, Anatomical basis of lingual hydrostatic deformation, *J. Exp. Biol.* 210, p. 4069-4082.
- Gomi H., Nozoe J., Dang J. & Honda K., 2006, A physiologically based model of perioral dynamics for various lip deformations in speech articulation, in J. Harrington & M. Tabain (eds.) *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press, New-York, p. 119-134.
- Hamann S., 2002, Retroflexion and retraction revised. *ZAS Working Papers in Linguistics* 28, 13-25.
- Hamann S. & Fuchs S., 2010, Retroflexion of voiced stops: Data from Dhao, Thulung, Afar and German, *Language and Speech* 53(2), p. 1-36
- Hashimoto K. & Suga S., 1986, Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue, *J. Acoust. Soc. Japan (E)* 7(1), p. 39-46.
- Kakita Y. & Fujimura O., 1977, Computational model of the tongue: A revised version, *J. Acoust. Soc. Am.* 62, S15(A)
- Kakita Y., Fujimura O. & Honda K., 1985, Computation of mapping from muscular contraction patterns to formant patterns in vowel space, in V.A. Fromkin (ed.), *Phonetic Linguistics*. Academic Press, Orlando, FL, p. 133-144.
- Kim K. & Gomi H., 2007, Model-based investigation of control and dynamics in human articulatory motion. *J System Design and Dynamics* 1, p. 558-569.
- Kiritani S., Miyawaki K., Fujimura O. & Miller J.E., 1976, A computational model of the tongue *Annual Bulletin of the Research Institute of Logopedics and Phoniatic* 10, Tokyo University, p. 243-251, URL: [http://www.umin.ac.jp/memorial/rilp-tokyo/R10/R10\\_243.pdf](http://www.umin.ac.jp/memorial/rilp-tokyo/R10/R10_243.pdf)
- Laboissière R., Ostry D. J. & Feldman A. G., 1996, The control of multi-muscle systems: human jaw and hyoid movements, *Biol. Cybern.* 74, p. 373-384.
- Ladefoged P. & Maddieson I., 1996, *The sounds of the world's languages*. Cambridge, MA: Blackwell.
- Liljencrants J. & Lindblom B.E.F., 1972, Numerical simulation of vowel quality systems: the role of perceptual contrast, *Language* 48, p. 839-862.
- Lindblom B., 1963, Spectrographic study of vowel reduction, *J. Acoust. Soc. Am.* 35, p. 1773-1781.
- Lindblom B., 1983, Economy of speech gestures, in P.F. MacNeilage (ed.), *The production of speech*, New York, Springer Verlag, p. 217-245.
- Lindblom B., 1984, Can the models of evolutionary biology be applied to phonetic problems? *Proceedings of the X<sup>th</sup> International Congress of Phonetic Sciences*, Utrecht, p. 67-81.
- Lindblom B., 1986, Phonetic universals in vowel systems, in J.J. Ohala (ed.). *Experimental Phonology*, New-York, Academic Press, p. 13-44.
- Løevenbrück H. & Perrier P., 1993, Vocalic reduction: prediction of acoustic and articulatory variabilities with invariant motor commands, *Proceedings of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany.
- Lucero J. & Munhall K., 1999, A model of facial biomechanics for speech production, *J. Acoust. Soc. Am.* 106(5), p. 2834-2842
- Lundberg A. & Stone M., 1999, Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data, *J. Acoust. Soc. Am.* 106(5), p. 2858-2867



- Ohala J.J., 1983, The origin of sound patterns in vocal tract constraints, in P. F. MacNeilage (ed.), *The production of speech*, New York, Springer Verlag, p. 189-216
- MacNeilage P. F., 1998, The frame/content theory of evolution of speech production, *Behavioural and Brain Sciences* 21, p. 499-511.
- McMahon T.A., 1984, *Muscles, Reflexes, and Locomotion*, Princeton University Press, Princeton, N.J.
- Miyawaki K., 1974, A study on the musculature of the human tongue: Observations on transparent preparations of serial sections, *Annual Bulletin of the Research Institute of Logopedics and Phoniatic* 8, Tokyo University, p. 23-50.
- Miyawaki K., Hirose H., Ushijima T. & Sawashima M., 1975, A preliminary report on the electromyographic study of the activity of lingual muscles, *Annual Bulletin of the Research Institute of Logopedics and Phoniatic* 9, Tokyo University, p. 91-106.
- Morimoto T., Inoue T., Nakamura T. & Kawamura Y., 1984, Frequency dependant modulation of rhythmic human jaw movements. *Journal of Dental Research* 68, p. 1310-1314
- Nazari M.A., Perrier P., Chabanas M. & Payan Y., 2010, Simulation of dynamic orofacial movements using a constitutive law varying with muscle activation, *Computer Methods in Biomechanics and Biomedical Engineering* 13(4), p. 469 - 482
- Netter F. H., 2004, *Atlas d'anatomie humaine*. Masson, Paris, France, Third edition.
- Ohala J.J., 1975, The temporal regulation of speech, in G. Fant & M. A. A. Tatham (eds.), *Auditory analysis and the perception of speech*, New York: Academic Press, p. 431-453.
- Padgett J., 2008, Glides, vowels, and features, *Lingua*, 118.
- Payan Y & Perrier P., 1997, Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis, *Speech Communication*, 22, p. 185-205.
- Perkell J. S., 1974, *A physiologically oriented model of tongue activity in speech production*, Unpublished Ph.D. thesis, Massachusetts Institute of Technology, Boston, USA
- Perrier P., Løvenbrück H. & Payan Y., 1996a, Control of tongue movements in speech: The Equilibrium Point hypothesis perspective, *J.Phonetics* 24, p. 53-75.
- Perrier P., Ostry D.J. & Laboissière R., 1996b, The equilibrium Point Hypothesis and its application to speech motor control, *J. Speech Hear. Res.* 39 (2), p.365-378.
- Perrier P, Payan Y, Zandipour M & Perkell J., 2003, Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study, *J. Acoust. Soc. Am.* 114(3), p. 77-83.
- Perrier P. & Ma L., 2008, Speech planning for V<sub>1</sub>CV<sub>2</sub> sequences: Influence of the planned sequence, *Proceedings of the 8<sup>th</sup> International Seminar on Speech Production (ISSP 2008)*, Université de Strasbourg, France, p. 69-72
- Piterman M., 2004, Chaos dans la modélisation des tissus mous, *Proceedings of the XXV<sup>th</sup> Journées d'Etude sur la Parole (JEP2004)*, Association Francophone de la Communication Parlée, p. 401-404.
- Prince A. & Smolensky P., 2004, *Optimality Theory: Constraint Interaction in Generative Grammar*, Oxford: Blackwell.

- Sanguineti V., Laboissière R. & Ostry D. J., 1998, A dynamic biomechanical model for neural control of speech production, *J. Acoust. Soc. Am.* 103(3), p. 1615-1627.
- Schiavone P., Boudou T., Promayon E., Perrier P. & Payan Y., 2008, A light sterilizable pipette device for the in vivo estimation of human soft tissues constitutive laws, *Proceedings of the 30<sup>th</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)*, Vancouver, Canada, p. 4298-4301.
- Schwartz J.-L., Boë L.-J., Vallée N. & Abry C., 1997, The Dispersion-Focalization Theory of vowel systems, *J. Phonetics* 25(3), p. 255-286.
- Schwartz J.L., Abry C., Boë L.J. & Cathiard M.-A., 2002, Phonology in a Theory of Perception-for-Action-Control, in J. Durand & B. Laks (eds.) *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, p. 255-286.
- Schwartz J.L., Boë L.J. & Abry C., 2007, Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAFF) principle in a Perception-for-Action-Control Theory (PACT), in M.J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology*, Oxford University Press, p. 104-124.
- Sifakis E, Selle A, Robinson-Mosher A & Fedkiw R., 2006, Simulating speech with a physics-based facial muscle model, in M.P. Cani, and J.O. Brien, editors, *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, p. 261-270.
- Stavness I., Hannam A.G., Lloyd J.E. & Fels S., 2010, Predicting muscle patterns for hemimandibulectomy models, *Computer Methods in Biomechanics and Biomedical Engineering* 13(4), p. 483-491
- Story B. H., 2005, A parametric model of the vocal tract area function for vowel and consonant simulation. *J. Acoust. Soc. Am.* 117(5), p. 3231-3254.
- Takemoto H., 2001, Morphological analysis of the human tongue musculature for three dimensional modelling, *J. Speech Lang. Hear. Res.* 44, p. 95-107.
- Terzopoulos D. & Waters K., 1993, Analysis and synthesis of facial Image sequences using physical and anatomical models, *IEEE Trans. Pattern. Anal. Mach. Intell.* 15, p. 569-579
- Turner M. J., Clough R.W., Martin H. C. & Topp, L. J., 1956, Stiffness and deflection analysis complex structures, *Journal of Aeronautical Science* 23, p. 805-823.
- Wilhelms-Tricarico R., 1995, Physiological modeling of speech production: methods for modeling soft-tissue articulators, *J. Acoust. Soc. Am.* 97(5), p. 3085-3098.
- Wilhelms-Tricarico R., 2000, Development of a tongue and mouth floor model for normalization and biomechanical modeling, in *Proceedings of the 5th speech production seminar and CREST Workshop on models of speech production* (Kloster Seeon, Bavaria), p. 141-148.