Why should speech control studies based on kinematics be considered with caution? Insights from a 2D biomechanical model of the tongue.

Yohan Payan & Pascal Perrier

Institut de la Communication Parlée - UPRESA CNRS 5009
INPG & Université Stendhal
46 Avenue Félix Viallet - F - 38031 Grenoble Cedex 01 - France
(payan; perrier)@cristal.icp.grenet.fr

ABSTRACT

A 2D biomechanical model of the tongue is used to simulate movement sequences and speech signals in Vowel-to-Vowel transitions. The analysis is focused on how central commands and biomechanics can interact and influence the physical speech signals. In particular, it is shown how complex velocity profiles can be explained by the biomechanics, and how the low-pass filtering effect of the biomechanics can give an account of the vocalic reduction phenomenon that is observed during speech production.

1. Introduction

This study aims to evaluate the impact of anatomical, morphological and biomechanical properties of the speech apparatus onto the kinematic properties of speech movement. Indeed many studies of speech motor control were based on features of the velocity profiles observed during articulatory displacements (Morasso, 1981; Soechting and Lacquaniti, 1981; Nelson, 1983; Ostry and Munhall, 1985; Munhall et al., 1985; Ostry and Flanagan, 1989; Adams et al., 1993). As already underscored in Perrier et al.'s. (1996a) paper with a physical model of the jaw, a major flaw of those studies lies in their incapability to make the difference between the kinematic properties arising from the way the articulators are controlled, and those strictly linked to the physical properties of the speech apparatus. To clarify the interactions of those two phenomena, a 2D biomechanical model of the tongue, controlled according to the *Equilibrium Point Hypothesis* of Feldman (1966), was set up. It was evaluated by generating articulatory trajectories in Vowel-to-Vowel sequences.

2. The 2D Tongue Model

2.a Basic Principles

In building the model, two main constraints were taken into account: a fair adequacy to the anatomical and mechanical structures of this articulator, and a limited computational complexity of the model, to permit a large number of simulations. From this perspective a major decision was to reduce the description of tongue structure to the midsagittal plane. Obviously, some important geometrical features of the tongue such as tongue grooving in [i] or lateral articulations, cannot be accounted for this way. However in spite of its simplicity, this approach is quite relevant, since it is in accordance with the main phonetic descriptions that classify speech sounds in terms of high-low/front-back positions of the tongue in the vocal tract. Thus it was considered to be reliable enough to assess our model for the main movements underlying the production of vowels. In addition, this choice is convenient, since the great majority of tongue movements measurement techniques (micro-beam, Electromagnetic Midsagittal Articulometer (EMA), or X-rays) are limited to the midsagittal plane.

Consequently the tongue model consists of a 2D Finite Element structure, which external shape of the rest configuration was adapted to tongue contours of a male speaker of French (Speaker PB), and was inserted in the external contours of the same vocal tract. Articulatory information about this speaker was extracted from an X-ray picture (Badin *et al.*, 1995), close to the production of a *schwa*.

2.b Muscle Description in the Model

As in other biomechanical tongue models published in the literature (Perkell, 1974, Kiritani et al., 1976; Hashimoto & Suga, 1986; Wilhelms-Tricarico, 1995), the global distribution of model’s nodes reflects the projection of the internal muscle structure into the sagittal plane. The number of elements was determined as the result of a compromise between anatomical accuracy and low computational costs. The great majority of the ten muscles or so that shape the tongue are in fact 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 relevant asymmetrical use of muscles in speech. 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. 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*. Muscles' insertions on bony parts (jaw and hyoid bone) were simulated by imposing “don't move” constraints to the corresponding nodes of the finite element lattice. At the bottom of the model (between the mental spine and the hyoid bone), the mylo-hyoid effect is modeled by a reaction force, that is applied to the corresponding nodes to limit the amplitude of the downward movements (Payan & Perrier, 1997).

2.c Low-Level Motor Control Mechanisms

According to Feldman's Equilibrium Point Hypothesis, movements are produced with constant rate shifts of the control variables. This determines a *virtual trajectory* in the control 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. Thus, the basic principles of the tongue model control for vowel production are as follows (Payan et al., 1995):

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

In addition, we propose that for a given phonetic context, the equilibrium target associated to a phoneme is unique. However, for the same phoneme, this target should *a priori* change, if the phonetic context varies, as the result of a planning process (Perrier *et al.*, 1996b; see also Guenther, 1995 for a mathematical formalism accounting for such a planning process).

The modeling of tongue muscle force generation has been directly inspired from the work carried out by Feldman and his colleagues (Feldman, 1966; Laboissière at al., 1996). The model assumes that afferent inputs related to muscle length and velocity are acting, together with descending central input, onto the α MNs depolarization to produce a global level of muscle force. Thus, muscle activation is represented by the difference between the actual muscle length and the recruitment threshold.

3. Simulations

The biomechanical tongue model and the hypotheses about its control in speech production were first evaluated through a comparison of the simulations with data collected on the speaker PB (Payan & Perrier, 1997).

3.a Reference Corpus

Basically, the corpus was designed to study V-V sequences, where tongue is the most relevant articulator. 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. Preliminary acoustic recordings were carried out, to verify that this bite-block did not actually induce an observable reorganization of the speech production strategy. During the experiment, the subject produced all possible French V-V sequences within two sets of vowels: the spread vowels [i, e, ε, a] and the rounded ones [y, o, u]. Articulatory data were collected with a five-transducers Electromagnetic Mid-sagittal Articulometer (AG100 system by Carstens Electronics). Three transducers were glued onto the tongue. The remaining transducers were glued onto the superior and inferior incisors.

3.b Interaction between Biomechanics and Kinematics

A comparison of the kinematic features measured from the data collected on the speaker, with those that are generated by the model is provided in Payan & Perrier (1997). The articulatory velocity profiles were at the center of this evaluation, since they were many times suggested to be relevant to the study of motor control. It was thus first demonstrated that our model of control, including the hypothesis of simple linear target-to-target trajectories in the control space, is able to generate realistic spatio-temporal patterns of tongue movements for the whole set of simulated Vowel-to-Vowel transitions. In addition, it was shown that, depending on the recruited muscles, a same temporal pattern of control variables can generate various velocity profiles, just as observed in real data. For instance, it was suggested that velocity profiles that display two velocity-peaks during the same target-to-target movement, should not necessarily be modeled as a sequence of sub-movements, unlike the hypotheses that are often proposed in the literature (see for instance Adams et al., 1993). Actually our results have suggested that it could simply be the result of morphological properties of the tongue (see Payan & Perrier, 1997, for details).

These simulations are a first indication that speech control studies based on kinematic features have to be considered with caution, as long as the underlying mechanical process is not properly taken into account. A second indication supporting this statement was provided by some attempts to simulate, with the biomechanical model, some aspects of the observed variability in speech physical signals.

3.c Speech Variability Prediction

Vowel reduction phenomena has been at the center of a large debate about the notion of target in speech (see for instance Lindblom, 1963; Gay, 1978 ; Pols and Van Son, 1993). In Perrier et al. (1996b), we have proposed a quantitative modeling of target-oriented speech production, to assess to which extent speech variability can be generated from invariant targets by controlling duration, context and speech style. This was carried out by using a functional dynamical model of the articulators, based on simple second-order principles.

Similar simulations were run with the more elaborated biomechanical tongue model. Figure 1 shows the prediction that was thus made for an [iai] sequence, of the variability in the articulatory and acoustical domains, when the hold duration of vowel [a] was reduced from

35 ms to 5 ms, while keeping constant the underlying target equilibrium positions. It is thus clearly confirmed that the low-pass filtering effect of the biomechanics can explain the reduction that is observed in speech data. In this perspective no change in the underlying target would thus be necessary to predict the vowel reduction phenomenon.

4. Conclusion

The simulations of VV transitions that were made with the model have shown to which extent morphological and dynamical properties of the speech apparatus are likely to influence the kinematic patterns measured on speech signals. From our point of view, a proper account of the biological properties of human speech production system, from the neurophysiology of the control to the mechanical properties of the peripheral apparatus, is highly recommended to understand from kinematic and acoustical data how speech production is controlled.

References

S.G. Adams, G. Weismer & R.D. Kent (1993), "Speaking Rate and Speech Movement Velocity Profiles", *Journal of Speech and Hearing Research*, Vol. 36, pp. 41-54.

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

A.G. Feldman (1966), "Functional Tuning of The Nervous System with Control of Movement or Maintenance of a Steady Posture – II Controllable Parameters of the Muscles", *Biophysics*, Vol. 11, pp. 565-578.

K. Hashimoto and S. Suga (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), Vol. 7 (1), pp. 39-46.

M. Hogan (1984), "An organising principle for a class of voluntary movements", Journal of Neuroscience, Vol. 4 (11), pp. 2745-2754.

T. Gay (1978). "Effects of speaking rate on vowel formant movements.", *Journal of the Acoustical Society of America*, Vol. 63, N°1, 223-230.

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

S. Kiritani, K. Miyawaki, & O. Fujimura (1976), " A Computational Model of the Tongue", *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, Vol. 10, pp. 243-252, University of Tokyo.

R. Laboissière, D.J. Ostry & A.G. Feldman (1996), "The Control of multi-muscle systems: human jaw and hyoid movements", *Biological Cybernetics*, Vol. 74(3), pp. 373-384.

B. Lindblom (1963). "Spectrographic study of vowel reduction"., *Journal of the Acoustical Society of America*, 35, 1773-1781.

P. Morasso (1981), "Spatial control of arm movements", *Experimental Brain Research*, Vol. 42, pp. 223-227.

K.G. Munhall, D.J. Ostry & A. Parush (1985), "Characteristics of velocity profiles of speech movements", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 11, pp. 457-474.

W.L. Nelson (1983), "Physical principles for economies of skilled movements", *Biological Cybernetics*, Vol. 46, pp. 135-147.

D.J. Ostry & K.G. Munhall (1985), "Control of rate and duration of speech movements", *Journal of the Acoustical Society of America*, Vol. 77, pp. 640-648.

D.J. Ostry & J.R. Flanagan (1989), "Human jaw movement in mastication and speech", *Archs oral Biol.*, Vol. 34 (9), pp. 685-693.

Y. Payan, P. Perrier & R. Laboissière (1995). "Simulation of tongue shape variations in the sagittal plane based on a control by the Equilibrium-Point hypothesis", *Proceedings of the 13rd International Congress of Phonetic Sciences*, Vol. 2, pp. 474-477. Stockholm, Sweden.

Y. Payan & P. Perrier (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, Vol. 22(1), July/August 1997

J.S. Perkell (1974), *A physiologically-oriented model of tongue activity in speech production*, Ph.D. Thesis. Boston: Massachussetts Institute of Technology.

P. Perrier, D.J. Ostry & R. Laboissière (1996a), "The equilibrium Point Hypothesis and its application to speech motor control", *Journal of Speech and Hearing Research*, Vol. 39 (2), pp. 365-377.

P. Perrier, H. Lœvenbruck & Y. Payan (1996b), "Control of tongue movements in speech: The Equilibrium Point hypothesis perspective", *Journal of Phonetics*, Vol. 24, pp. 53-75.

L.C.W. Pols & R.J.J.H. Van Son (1993). "Acoustics and perception of dynamic vowel segments". *Speech Communication*,. 13, 135-147.

J.F. Soechting & F. Lacquaniti (1981), "Invariant characteristics of pointing movement in man", *Journal of Neuroscience*, Vol. 1, pp. 710-720.

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

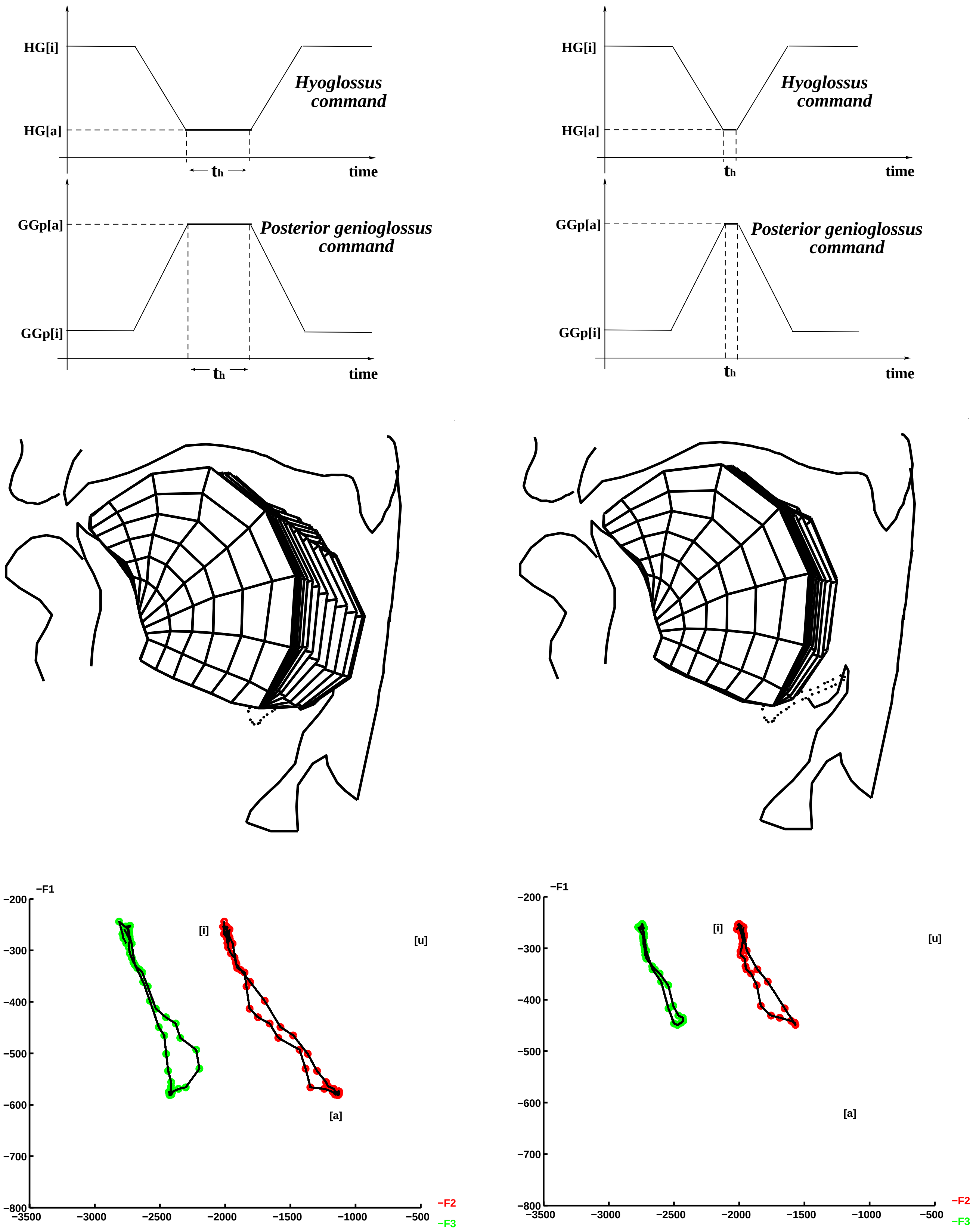

Figure 1: Simulation of vowel reduction in the [iai] sequence when [a] hold duration is reduced from 35ms (Normal) to 5ms (Reduced), without any change in the target equilibrium positions. Left handside:Normal; Right handside: Reduced. Top Panel: Variation of the control variables over time. Middlepanel: Simulated tongue movements. Low panel: Trajectories in F1/F2 and F2/F3 planes (in Hz).