MODELLING OF SPEECH MOTOR CONTROL AND ARTICULATORY TRAJECTORIES

P. Perrier, R. Laboissière & L. Eck.

Institut de la Communication Parlée - URA CNRS n° 368
INPG-ENSERG,Grenoble, France.

ABSTRACT

In order to study motor control strategies in speech production, we propose to simulate the dynamical behaviour of speech articulators with a stiffness commanded distributed second order model, where the set agonist/antagonist is commanded as a whole. For [CVCV] utterances, inversion from the jaw movements to the corresponding stiffness commands is proposed using a guided algorithm of error backpropagation. We focused the analysis of our results on the ability of our model to predict target undershoot, and to detect hypo- and hyper-articulation strategies used by two speakers.

1. INTRODUCTION

The so-called *Equilibrium Point Hypothesis*, introduced by Asatryan & Feldman ([2]), suggests that skilled movements correspond to shifts in the equilibrium state of the motor system. In this framework, two major, quite different kinds of modeling have been developped: the λ *model* proposed by Feldman and his co-workers (see [2], [6] and [7]) and α *model*, first proposed by Bizzi [3] (see also [4]). According the former the commanded variable is the *recruitment threshold* of the used muscle, whereas in the latter the muscle stiffness (corresponding to the muscle activation) is controlled. Both approaches yielded very appealing results for jaw or multi-joint arm movements ([4] and [7]). But in his clarification article [6] Feldman develops his argumentation against Bizzi's α *model*: this latter can actually neither explain movements occurring with a constant muscle activation level nor the absence of movement for a certain kind of variation of this activation; moreover stiffness cannot be centrally commanded since, due to afferent signals, this variable depends on the length of the muscle and then varies during the movement.

From our point of view, if the *agonist and antagonist muscles* are considered as a whole, the concept of equilibrium point defined as an *equilibrium between agonist and antagonist stiffnesses* is functionnally very appealing. Thus, in spite of everything, we proposed [10] to use a stiffness commanded distributed second order model for the set agonist/antagonist muscles considered as a specifically commanded whole; the advantage of this global modeling lies in the fact that it overcomes the main critics of Feldman against the stiffness model (see further). In order to test the validity of such an approach, we propose here to confront our model with data on jaw movements in the production of CVCV sequences. Inversion using a error back propagation algorithm is studied in order to infer the stiffness commands which allow the generation of suitable trajectories. The results are analysed in regard to the control strategies proposed by this technique for certain kinds of speech production.

2. OUR SECOND ORDER MODEL

According to the kinematics characteristics of skilled movements presented by Nelson [9], our model [10] (see fig.1) consists in a couple of springs, one for the agonist set and another for the antagonist one; these springs are linked by a material point, whose mass (m) is normalized to 1. The displacements of this point correspond to shifts from an equilibrium point of the system to another. This latter is called *target* of the movement. The successive targets (or equilibrium points) are determined by the ratio (η) between the stiffness (k_1 and k_2) of the two springs. These mechanical targets are directly linked to the underlying phonetic targets of the sequence: each vowel and each consonant correspond to a specific value of the ratio η.

Both springs have the same rest length x_0. When the equilibrium point of the system is shifted, an unidirectional movement of the material point occur. Let x be the spatial variable in the direction of the movement; the dynamical equation describing the system is then:

$$\frac{\partial^2 x}{\partial t^2} = -f.\frac{\partial x}{\partial t} - (k_1+k_2).x - (k_1-k_2).x_0 \quad (1)$$

It is, of course, quite easy to notice from this equation that, given f, the values of k_1 and k_2 determine completely the trajectory. Because we consider the agonist and antagonist sets as a whole, we propose to command the model with two variables which act simultaneously on these sets:

- the *stiffness ratio* η which determine the equilibrium position of the target;
- the *cocontraction K*, corresponding to the global activation k_1+k_2 of the set agonist/antagonist.

The major critics of Feldman [6] against the stiffness model don't apply to our approach: it is actually obvious that movement can occur without modification of the cocontraction level (with a reciprocal variation of k_1 and k_2), and that this level can be modified without change in the resulting stiffness ratio η and therefore without movement. Moreover the cocontraction level is not dependent on the length of each spring and can therefore be centrally commanded. This holds if we suppose a symetrical modeling of the agonist and antagonist sets, which would induce a reciprocal lengthening/shortening on each of them.

The commands η and K vary theoretically by step between targets. However, in order to propose more realistic variations of the commands, we have smoothed the abrupt transitions of these signals by filtering them with a critical second order filter (τ=80 ms). The duration of each step is explicitely commanded.

3. THE INVERSION APPROACH

Equation (1) describes the dynamics of the model, where K and η are the inputs and x is the output. The goal of the inversion procedure is to infer the time-varying functions K(t) and η(t) that generate the actual jaw displacement x(t). Since equation (1) does not have constant coefficients, it is quite hard to derive an analytical solution to the general inversion problem. We applied then an iterative optimization procedure, essentially a gradient-descent technique, where we minimize a cost functional given by the squared error between the actual and the model output signals integrated over the time interval of interest. We carry this optimization over the space of possible functions K(t) and η(t), with the constraints described in section 2.

The gradient of the cost functional can be obtained using the calculus of variations, but for a discrete version of (1) (see [5]) we obtain a formulation close to the *error backpropagation through time* [11]. Without truncation in time, this method give an exact gradient and the error cost tends asymptotically to a local minimum through the iterations. With good guesses for the initial state and some interactive control during the process (e.g. alternating the optimization of the duration and amplitude of the commands) we get reasonable results, like those shown in the following section.

4. RESULTS - DISCUSSION

4.1. Description of the corpus

The corpus consists of the utterance [zɛzzɑ] in Tunisian Arabic (what means: "he rewarded"). It is pronounced within a carrier sentence at two different speech rates ("normal" and "as fast as possible") and by two different native speakers. The movements of the jaw are considered here to be pertinent enough for a reliable description of the production strategies. They were recorded with a mandibular kinesiograph (K5AR), and sampled at 160 Hz (for more details see [1])

4.2.Undershoot phenomenon in the inter-consonantal vowel

In the following we denote by $\Delta\eta$ the amplitude of the variation of η, and by $\Delta\tau$ the temporal percentage of the vocalic command within the total duration of the vocalic plus consonantal commands.

We tried to fit the output of our model to the jaw data from one speaker at the two rates. First of all, *the level of concontraction K of the model was held constant*. Fig.2 shows the corresponding results :

- at normal rate, the spatial positions corresponding to the mechanical equilibrium points (called *ideal targets*) are reached for both consonants [z] and [zz], but a slight undershoot occurs for the vowel [ɛ]; $\Delta\eta$ is 0.49, and $\Delta\tau$ is 33%.
- at fast rate, the ideal targets are reached for the consonants, and we observe a very clear undershoot for the vowel; $\Delta\eta$ is 0.62, and $\Delta\tau$ is 34% .

At first glance, these results are satisfying:

through our inversion, the well-known *vocalic reduction* phenomenon due to speech rate increasing (see [8]) stands out. However the underlying command strategy here proposed seems to be unrealistic: $\Delta\eta$ increases in the case of vocalic reduction. This would mean that the speaker point to a further target to minimize the undershoot!!!

We adopt then the same fitting approach but with simultaneous optimization of K and η. The results (Fig.3) are more satisfying:

- at normal rate, K remains approximately the same as above for the consonant, but increases for the vowel production; all ideal targets are reached, $\Delta\eta$ is 0.41 and $\Delta\tau$ is 41%; it seems then that in order to prevent any influence of the consonantal context on the vowel, the speaker makes a particular effort for the vocalic articulation, corresponding to the increase of K.
- at fast rate, we observe a clear undershoot in the production of the vowel; we notice a reduction of the vocalic duration ($\Delta\tau$=33%) and a decrease of the cocontraction level for the vowel production. The vocalic reduction could thus be explained through a credible strategy: the instruction "speak as fast as possible" induces in the speaker *a decrease in his articulation effort*, corresponding to a decrease of the cocontraction level for the vowel production. From this point of view this second inversion is very interesting. However we observe again an increase of $\Delta\eta$, whose value is 0.53. This can be explained by the fact that too many parameters (K, η, and the durations of each command step) have to be optimized at the same time for this simulation. In order to get a better inversion, we have to propose constraints on the respective evolutions of these parameters; for example, the contraint "$\Delta\eta$ must be the same for normal and fast rates" would consolidate the above assumed strategy of our speaker for vocalic reduction.

4.3. Hypo- and hyperspeech strategies

Our further point is to compare the production strategies used by two different native speakers for the same utterance. Fig.4 depicts the results of the inversion in the same conditions as just above. For both normal and fast rates, all ideal targets are reached for our second speaker: the cocontraction level increases strongly for the vowel production, and particularly at fast rate; this speaker seems to increase his articulation effort when speech rate increases. This assumption corresponds to an audible characteristic of the speech signal: at fast rate this speaker cries out!!! We think so that our model provides a good tool to detect the phenomenon of hypo/hyperarticulation, as proposed by Lindblom ([8]), from the articulatory data. At fast rate, the first speaker tends to hypoarticulate whereas the second one tends to hyperarticulate.

5. CONCLUSION

By means of an error backpropagation technique we were able to fit available data on jaw movement to the output of a model consisting of agonist/antagonist pair of springs. The controlled variables in this model are the stiffnesses of the springs taken as a whole. In spite of Feldman's interesting critics against stiffness control for skilled movements, we showed that our model can explain known phenomena in speech production, namely vowel reduction and hypo/hyperarticulation strategies.

ACKNOWLEDGEMENTS

To Christian ABRY for his incitation on working in this direction, and to Jomaa MOUNIR for the data collection on jaw movements.
The second author was supported by a scolarship from the National Research Council (CNPq), Brazil, under FIAS-CNPq file number 92.0208/88.6. He is also with the Aeronautics Technology Institute (ITA), São José dos Campos, Brazil.

REFERENCES

[1] ABRY C., PERRIER P. & JOMAA M. (1990), "Premières modélisations du timing des pics de vitesse de la mandibule" *Proceedings of the 18th J.E.P.*, Société Française d'Acoustique, 99-102.
[2] ASATRYAN D.G & FELDMAN A.G. (1984), "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. 1.Mecanographic analysis of the work of the joint on execution of a postural task," *Biophysics*, *10*, 925-935.
[3] BIZZI E. (1980), "Central and peripheral mechanism in motor control," in *Tutorials in motor behavior*, G.E. Stelmach & J. Requin (eds.), Amsterdam: North-Holland, 131-144.
[4] COOKE J.D. (1980), "The organization of simple skilled movements," in *Tutorials in motor behavior*, G.E. Stelmach & J. Requin (eds.), Amsterdam: North-Holland,
[5] ECK L. (1990), "*Modélisation des gestes articulatoires*," Mémoire de DEA, Institut National Polytechnique de Grenoble, France.

[6] FELDMAN A.G. (1986), "Once more on the equilibrium-point hypothesis (λ model) for motor control," *Journal of Motor Behavior*, *18*, *1*, 17-54.
[7] FLANAGAN J.R., OSTRY D.J. & FELDMAN A.G. (1990), "Control of human jaw and multi-joint arm movements," in *Cerebral Control of Speech and Limb Movements*, G. Hammond (ed.), London: Springer-Verlag.
[8] LINDBLOM B.(1990], "Explaining phonetic variation: a sketch of the H&H theory,", in *Speech Production and Modelling*," W.J. Hardcastle & A. Marchal (eds.), London: Kluwer Academic Publishers.
[9] NELSON W.L. (1983), "Physical principles for economies of skilled movements," *Biol. Cybern.*,1-9.
[10] PERRIER, P., ABRY, C. & KELLER E. (1989), "Vers une modélisation des mouvements du dos de la langue," *J. Acoustique*, *2*, 69-77.
[11] WILLIAMS R.J. & ZIPSER D. (to appear), "Gradient-based learning algorithms for recurrent connectionist networks" in *Backpropagation: Theory, Architectures and Applications*, Y. Chauvin & D.E. Rumelhart (Eds.), Hillsdale, NJ: Lawrence Erlbaum.

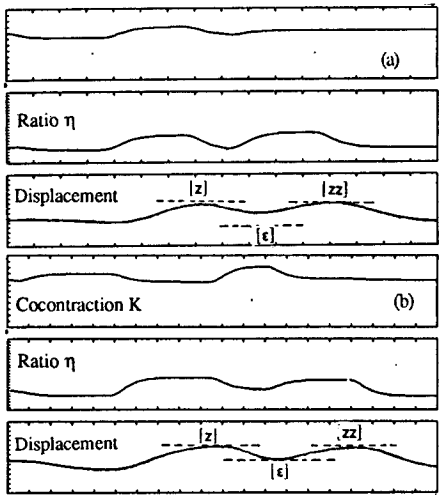

Figure 3: Stiffness commands and jaw displacement obtained for the first speaker; for comments see figure 2.

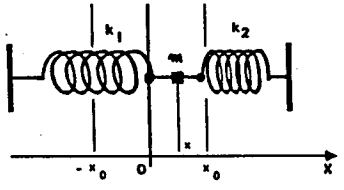

Figure 1: The distributed second order model.

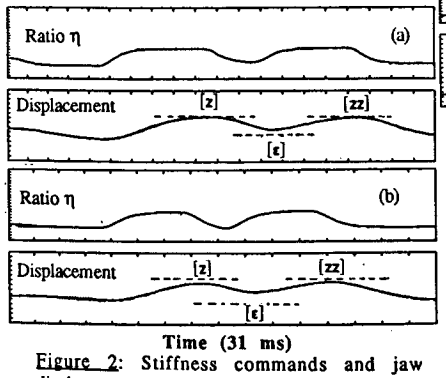

Figure 2: Stiffness commands and jaw displacement obtained for a constant cocontraction level; as regards the displacement, at the scale of the figure, there is no perceptible differences between model output and data; the dotted segments on the displacement curve correspond to the different infered *ideal targets* (see text); (a)=normal rate, (b)=fast rate.

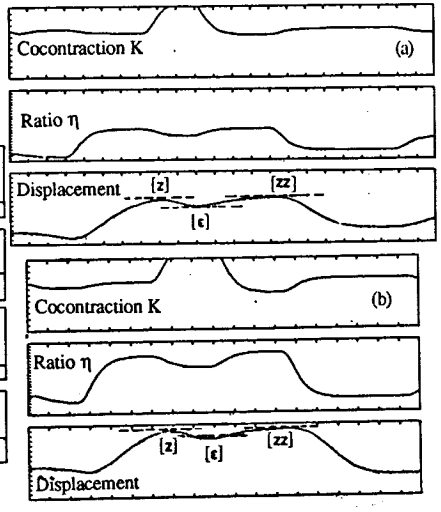

Figure 4: Stiffness commands and jaw displacement obtained for the second speaker; for comments see figure 2.