EQUILIBRIUM POINT HYPOTHESIS AND ARTICULATORY TARGETS IN SPEECH: A DESCRIPTION FROM SIMULATIONS OF EMPIRICAL DATA USING A BIOMECHANICAL MODEL OF THE JAW

LŒVENBRUCK Hélène[1], *PERRIER Pascal*[1] *& OSTRY David J.*[2]
[1]*Institut de la Communication Parlée, URA CNRS 368, INPG & Université Stendhal*
46, avenue Félix Viallet, 38031 GRENOBLE Cedex 01, France
[2]*Department of Psychology, Mc Gill Unversity,*
1205 Dr Penfield Avenue, MONTREAL Qc, Canada H3A 1B1

ABSTRACT

A previous study based on a simple model of speech articulator motion supported the idea that articulatory trajectories in vowel sequences could be produced from a succession of targets defined in terms of invariant equilibrium points. The prosodic variability was accounted for by modification in the timing of the commands and the level of force involved. These findings are compared with those obtained using a more sophisticated biomechanical model of the jaw.

INTRODUCTION

Controversies surrounding the notion of targets in speech production have centred on their nature (see eg [1] for a short review), on their timing ([2] *vs* [3]) and even on their existence [4].

The Equilibrium Point Hypothesis (EP Hypothesis) proposed by Feldman [5] for the control of limb movement provides some additional insight into this debate. According to this model, movements arise from changes in posture. In this context, speech targets may be associated with specific postures of the articulators and successive postures could correspond to a representation of the articulatory task at the level of control.

The notion of articulatory targets in speech production has been used in the debate on speech invariance and variability to support the idea that for a given phoneme, independent of the phonetic context, each articulator tends to approach a single position [6]. This idea that articulatory movements are intended towards spatial positions is related to MacNeilage's proposals [7].

The EP Hypothesis provides a neurophysiologically based account of how target positions can be specified [8]. In a previous study a simple second order model of the articulators, controlled in terms of equilibrium shifts, was used to examine how articulatory targets may be specified and how they may be related to an invariant linguistic task in different prosodic conditions [9].

In this paper, hypotheses based on simulations which used this simplified model are tested using a more sophisticated biomechanical model of the jaw [10].

STARTING HYPOTHESES

Lœvenbruck & Perrier [11] showed that the EP hypothesis can shed light on the control of vowel reduction. A second order model, consisting of two springs which simulated agonist and antagonist muscle groups was used to provide a simple way of controlling the gesture's dynamic parameters: the stiffness ratio specified the position of the intended spatial targets, whereas the cocontraction level (sum of the stiffnesses) and the timing of the commands determined the dynamic behaviour of the model.

The [iai] sequence was studied in the French sentence "Il y a immédiatement". The sequence was tested in three different conditions: (1) slow speech rate and stressed [a]; (2) slow rate and unstressed [a]; (3) fast rate and stressed [a]. Data from one French native speaker were analyzed. Articulatory trajectories were inferred from the acoustic signal by an inversion procedure involving an articulatory model. Condition (1) was supposed to be the "ideal" one, in the sense that the observed articulatory positions for [i], [a] and [i] correspond to the intended articulatory targets. In both the other cases, the movement extent was reduced.

Lœvenbruck & Perrier showed that the three different articulatory trajectories could be generated using the same successive intended articulatory targets for [i], [a] and [i]. The differences between the trajectories in the three conditions could be simulated by modifying the cocontraction level and the timing of target shift. In order for the model to replicate the kinematic patterns of stressed and unstressed vowels produced at slow speech rate it was necessary that cocontraction (and thus total force) was greater for the stressed vowel; no differences were observed in the timing of equilibrium shift. In contrast, the main difference between slow and fast stressed conditions was the timing of the shift; the cocontraction levels were almost identical. In the present paper, these findings serve as initial hypotheses for tests carried out with a more sophisticated model. Specifically, the timing of the commands should remain fairly constant at a given speech rate; movements for stressed vowels should involve greater total force than for unstressed vowels.

SIMULATIONS WITH A BIOMECHANICAL MODEL OF THE JAW

The jaw model

The model proposed by Laboissière et al.[10] has seven muscles (or muscle groups) and four kinematic degrees of freedom. Besides a more elaborate biomechanical formulation, this model has the advantage over the one used in the previous work of including a fuller account of the neurophysiological control mechanism. The essential control variables are independent changes in the membrane potentials of motoneurones (MNs) which establish a threshold muscle length (λ) at which the recruitment of MNs begins. Muscle activation and hence force vary in relation to the difference between the actual and the threshold muscle lengths and the rate of muscle length change. Thus, by shifting λ through changes to the central facilitation of MNs, the system can produce movement to a new equilibrium position.

In the jaw model, movements are not controlled directly in terms of commands to individual muscles. Rather control signals, which are based on different combinations of λs, are organised at the level of the system's kinematic degrees of freedom. This enables production of jaw rotation, horizontal jaw translation, vertical hyoid translation and horizontal hyoid translation. The level of cocontraction is also controlled by specifying the global level of force involved in the movement. These control signals may act alone or in combination.

Methods

The corpus was the same as the one used for the previous work. A different native speaker of French was tested.

Horizontal jaw translation and rotation in the midsagittal plane were tracked using the Optotrak system which captures the light emitted by infrared emitting diodes (IReds). An acrylic dental appliance accurately fitting the lower teeth was built from the dental impression of the subject. A lightweight but rigid dental wire was attached to the front of the appliance and was shaped so that its two ends came out of the mouth horizontally at the corner of the lips. A total of five IREDs were attached to bamboo sticks which were glued to the dental wire. These IREDs were used to track the motion of the jaw. An additional six IREDS were attached to a head mounted acrylic frame and were used to correct for head motion. IRED positions were sampled at 100 Hz and low pass filtered using a Butterworth filter. The acoustic signal was simultaneously recorded and sampled at 10 kHz. The orientation angles and positions which characterise the motion of the jaw were reconstructed from the IRED motions.

The [iai] sequence was extracted from the entire sentence using the Vocalic Voiced Onset and Vocalic Voiced Termination criteria [12].

Simulations

Only rotation of the jaw was considered because jaw rotation is the most relevant articulatory variable in the [i-a] transition, in the sense that the transition is characterised largely in terms of differences in jaw opening angle.

The same strategy as in the previous work was used, *ie* inferring the intended targets from the articulatory signal observed under the slow and stressed condition. We specifically tested the idea that stressed movements were associated with a high total force or high cocontraction level and changes in speaking rate were associated with changes in the duration of the command.

The results of the simulation for the slow and stressed condition are presented in figure 1, where the data and the simulated trajectories are plotted in dotted and solid lines respectively. As can be seen the fit to the data is rather good. Note that the actual position for the second [i] is influenced by the incoming context which is not taken into account in our simulations. The discrepancy between the data and the simulation for [a] is due to the dynamic coupling of sagittal plane rotation and horizontal translation. The actual [a] position slightly undershoots the target defined by the intended equilibrium position.

The fast and stressed condition is simulated by a reduction in the control signal underlying [a]. The amount of force is the same as in the previous condition. The fit here is likewise relatively good (figure 2).

Finally, the slow and unstressed condition was simulated by setting the global force to its minimum possible value and by reducing the equilibrium shift rate. The decrease in shift rate effectively eliminates a stationary position of [a]. Under these conditions, one may observe a clear reduction (-3 deg.) in the amplitude of the simulated movement. Under these conditions a suitable fit to the data was not possible (figure 3).

CONCLUSION

The simulation obtained for the fast and stressed condition could be obtained in a manner consistent with our initial hypotheses. A simple reduction of the hold duration for the vowel [a] produces a target undershoot comparable to that observed in the empirical data. The high level of global force as well as the relatively fast transition rate enable the simulation of the sharp transition which is observed. Moreover the reduction of movement amplitude obtained by decreasing the global force in the simulation of the slow and unstressed condition, confirms the role of the cocontraction level as an efficient parameter for dynamic control.

However reducing the force and the transition rate were not sufficient to produce an undershoot comparable to that observed empirically. This argues therefore against our starting hypothesis that articulatory targets remain the same independent of stress. A better fit to the empirical trajectory was obtained by reducing the amplitude of the equilibrium shift from [i] to [a] (figure 4). This can be interpreted in two ways:

(1) a change in stress corresponds to a change in the intended articulatory target.

(2) a decrease in the equilibrium shift rate induces an undershoot at the level of the control variables.

The absence of stationary target position for [a] in the slow and stressed condition favours the second hypothesis. However, further tests should be carried out, by comparing, for example, the articulatory patterns for [ɛ] and [a], which are acoustically similar under conditions of vowel reduction.

ACKNOWLEDGEMENT

This work is supported by the Esprit B.R. Project n° 6975, Speech Maps, by the cooperation France-Québec (Projet n° 07-01-92) and by the NIH grant DC-00594 from the National Institute of Deafness and Other Communication Disorders.

REFERENCES

[1] MacNeilage P. (1980). Distinctive properties of speech motor control. In G.E. Stelmach and J. Requin (eds.)*Tutorials in motor behavior* ,607-621. Amsterdam, The Netherlands: North Holland publishing company.

[2] Lindblom B., Lubker B, Branderu P. and Holmgren K. (1987). The concept of target and speech timing, 161-182, Channon R. and Shockey L. (eds). Honor of Isle Lehiste, Foris: Dordrecht, Holland.

[3] Fowler C.A. (1980). Coarticulation and theories of intrinsic timing. *J. Phonetics*, 8, 113-133.

[4] Pols C. W. & Van Son R.J.J.H. (1993). Acoustic and perception of dynamic vowel segments. *Speech Comm.* 13, 135-147.

[5] 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.

[6] Lindblom B. (1963). Spectrographic study of vowel reduction. *J. Ac. Soc. Am.* 35, 1773-1781.

[7] MacNeilage P. (1970). Motor control of serial ordering of speech. *Psy. Rev. 77, 182-196.*

[8] Perrier P. & Ostry D.J. (1994), "Dynamic modelling and control of speech articulators. Application to vowel reduction." In Keller E. (Ed.), *Fundamentals in Speech Synthesis and Speech Recognition* ,231-251. London, U.K. : J. Wiley and Son.

[9] Perrier P., Lœvenbruck H. & Payan Y. (submitted), "Control of tongue movements in speech: The Equilibrium Point hypothesis perspective.", *J. Phonetics.*

[10] Laboissière R., Ostry D.J. & Feldman A.G. (submitted). The Control of Human Jaw and Hyoid Movement. *J. of Neurophysiology.*

[11] Lœvenbruck H. & Perrier P. (1993). Vocalic reduction: prediction of acoustic and articulatory variabilities with invariant motor commands. *Acts of the 4th European Conference on Speech Communication and Technology* , 85-88. Berlin, RFA.

[12] Abry C , Benoit C, Boë L.J. & Sock R. (1985). Un choix d'événements pour l'organisation temporelle du signal de parole. Actes des 14èmes JEP Paris, 133-137.

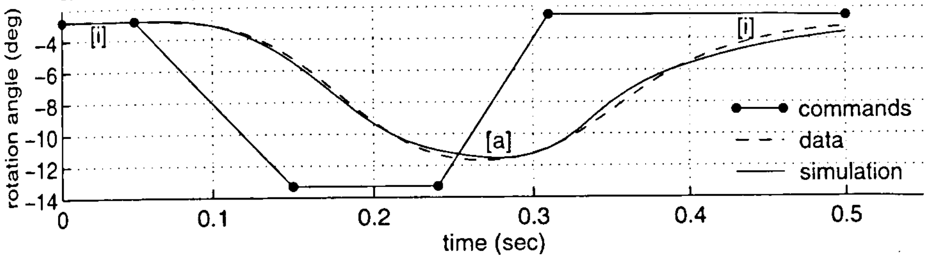

Figure 1 : Simulation under the slow and stressed condition. The equilibrium targets were E(i1) = -2.8 deg for the first [i], E(a) = -13.3deg for [a] and E(i2) = -2.5deg. for the last [i]. The level of force (computed for the first [i]) was F = 78.4N, the hold time for the first [i] was Thold1 = 0.05s, the [i-a] transition time was Tt1 = 0.1s, the hold time for [a] was Thold2 = 0.09s and the [a-i] transition time was Tt2 = 0.07s.

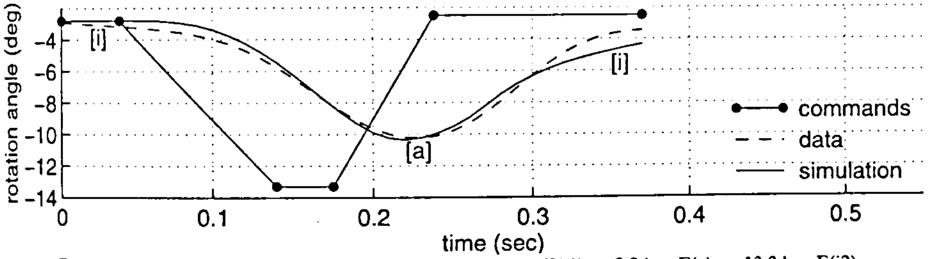

Figure 2 : Simulation under the fast and stressed condition: E(i1) = -2.8deg, E(a) = -13.3deg, E(i2) = -2.5deg., F = 78.4N, Thold1 = 0.04s, Tt1 = 0.1s, Thold2 = 0.035s and Tt2 = 0.063s.

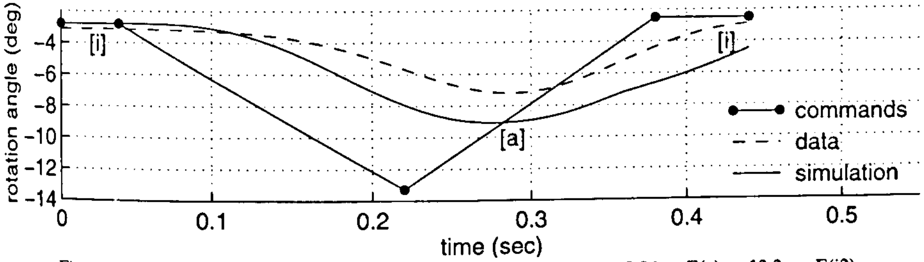

Figure 3 : Simulation under the slow and unstressed condition: E(i1) = -2.8deg, E(a) = -13.3 eg, E(i2) = -2.5deg., F = 10.6N, Thold1 = 0.04s, Tt1 = 0.179s, Thold2 = 0.001s and Tt2 = 0.16s.

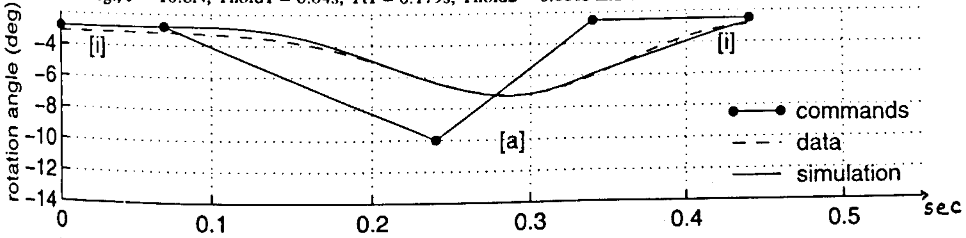

Figure 4 : Simulation under the slow and unstressed condition: E(i1) = -2.8deg, E(a) = -10deg, E(i2) = -2.5deg., F = 10.6N, Thold1 = 0.07s, Tt1 = 0.169s, Thold2 = 0.0001s and Tt2 = 0.1s.