1136

CONTROL OF COARTICULATORY PATTERNS OF TONGUE AND JAW MOVEMENT IN SPEECH

David J. Ostry[1,2], Douglas M. Shiller[1], Paul L. Gribble[1] and Vincent L. Gracco[2]
[1]McGill University, Montreal, Quebec, [2]Haskins Laboratories, New Haven, CT.

ABSTRACT

Orofacial movements arise from the interaction of neural control signals with mechanical and geometrical properties of vocal tract structures. We have previously shown that in the case of jaw movement, patterns of coarticulation may arise as a consequence of muscle properties and dynamics rather than characteristics of the neural code. This finding suggests that at the level of individual articulators the nervous system may take limited account of orofacial dynamics in planning movements. In the case of multi-articulator motion the mechanical interaction of articulators may necessitate a more complex pattern of control. In order to assess the characteristics of control signals that underlie inter-articulator patterning, we describe the kinematic patterns of coarticulation of tongue and jaw movements in speech. The empirically observed kinematic patterns are compared with the results of simulation studies that use a simplified model of tongue and jaw motion.

1. INTRODUCTION

Coarticulation in speech production is a phenomenon in which the articulator movement for a given consonant or vowel varies systematically with the surrounding movements. Coarticulation has been reported in a number of speech articulators and various experimental contexts. While a range of accounts of speech coarticulation have been proposed, two views have dominated the debate: context-invariant (superposition) models which involve the blending of the influences of invariant central commands and context-sensitive (look-ahead) models which suggest anticipatory adjustments in the central commands.

The major difference in these two competing views is the extent to which the kinematic variations due to context sensitivity are directly coded in neural signals. In the context invariant models, coarticulation effects "fall out" of the interaction between adjacent segments while in the context-sensitive approach the nervous system explicitly plans the different trajectories taking context into account. A fundamental difficulty in addressing these two competing explanations is that in order to infer control strategies from kinematic data alone one must know what aspects of the kinematic variation result from non-neural factors such as muscle mechanical properties and dynamics. Without explicit models of the speech articulators, the kinematic patterns which are attributable to central control cannot be distinguished from those which arise due to dynamics and are not represented in the underlying control signals.

We have recently shown, using a physiological model of jaw and hyoid motion [3], that some of the systematic variations associated with anticipatory and carryover coarticulation may result from the biomechanical properties of the vocal tract and not from the control input [4]. A more recent version of the model includes the tongue and associated inter-articulator mechanical interactions [6].

2. MODELS

Our models are based on the equilibrium point (EP) hypothesis of motor control [1]. The models include simulated neural signals and reflexes, muscle mechanics, realistic musculo-skeletal geometry, and articulator dynamics. The muscle model is variant of the Zajac [7] formulation and includes the dependence of force on muscle length and velocity, reflex damping and graded force development due to calcium kinetics (Figure 1).

The EP hypothesis proposes that movements result from shifts in the equilibrium state of the motor system. According to the model, force develops in proportion to the difference between the actual muscle length and a centrally specified threshold length for motoneurone recruitment, λ. Thus, by shifting muscle λs the system may move to a new equilibrium position.

In simulations with the model, we have shown that the smooth empirical patterns of jaw motion in speech can be achieved using constant rate shifts in the underlying equilibrium position of the jaw. We have also shown that the observed kinematic patterns of coarticulation can be simulated using constant rate equilibrium shifts. In order to assess the extent to which muscle properties and dynamics contribute to the observed coarticulatory patterns, we held the modeled central control signals fixed so that they take no account of context. Thus at a control level there is no planned coartic-

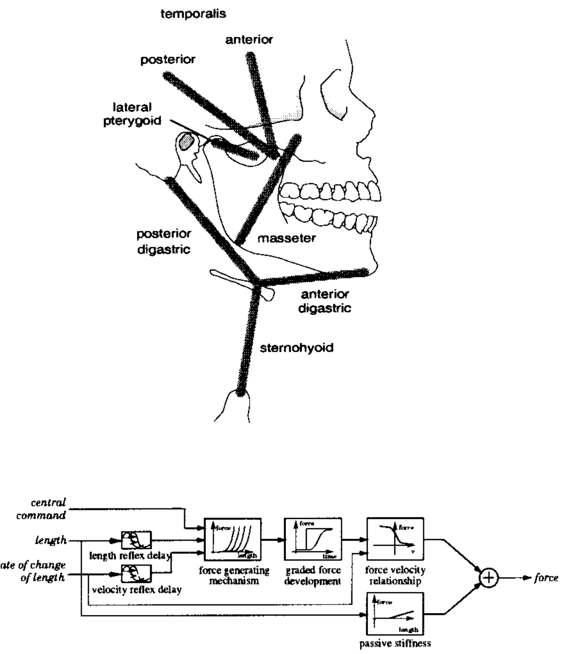

Figure 1: Top: Schematic of the modeled muscle groups and their attachments to the jaw and hyoid bone. Bottom: Muscle mechanical model.

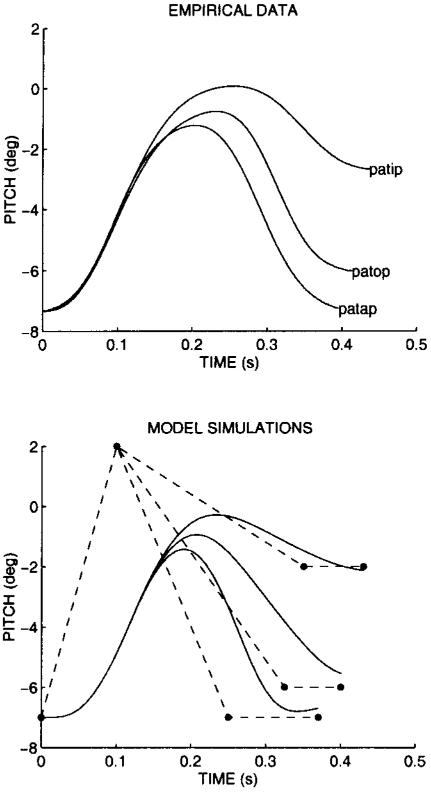

Figure 2: Comparison of empirical data and model predictions for anticipatory coarticulation.

ulation. We then examined the simulated kinematic patterns and found that they matched the empirically observed patterns of coarticulation. A summary of the obtained empirical and modeling results is given in Figure 2.

In the present experiment we extend our earlier experimental work to examine tongue and jaw motion in speech. We address the origins of coarticulatory variations by comparing the results of empirical and modeling studies. Initial results based on a simplified model of the tongue-jaw system are consistent with our previous findings suggesting that even when no account is taken of context at the level of central control, kinematic patterns of the tongue and jaw vary in amplitude, duration and timing in a manner similar to that observed empirically.

3. METHODS

Sagittal-plane movements of the lips, tongue and jaw were obtained using an electromagnetic transduction device. The prototype of this system has been described by Perkell et al. [5]. Receivers were placed in the mid-sagittal plane on the nose, the maxilla, the upper and lower lips, the mandible and the tongue. Two mandibular receivers were used, spaced approximately 1 cm apart, in order to decompose jaw motion into its rotational and translational components. Three receivers were placed on the tongue, approximately 1.5-2.0 cm apart, with the front-most tongue receiver placed approximately 1.5 cm back from the tongue tip.

Tongue and jaw kinematics were recorded during the repetitive production of *VCV* sequences in which the movement amplitudes for the initial and final vowels were varied. The focus of the study was the manner in which the movement amplitude and duration of the tongue and jaw raising movement associated with the transition between the initial vowel and consonant were varied as a function of the final vowel (anticipatory coarticulation).

The utterances were composed of the vowels *a*, *o* and *i* and consonants *k*, *t* and *s*. The vowels were chosen to produce differences in tongue and jaw movement amplitude. The different consonants varied the position and possible precision requirements of consonant articulation. The utterances were embedded between flanking consonants *p* to produce speech-like sequences such as *pakap*. All combinations of three initial vowels, three consonants and three final vowels were tested. Subjects were instructed to stress the first vowel and to produce the sequences repetitively at a

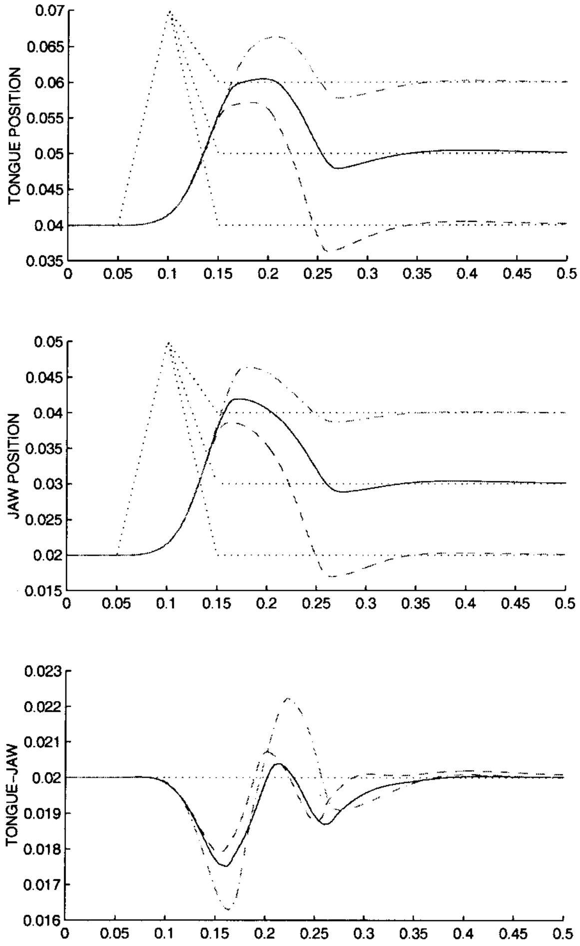

Figure 3: Top two panels: Predicted patterns of tongue and jaw vertical position (smooth curves) and underlying control signals (dots). Bottom panel: Predicted difference between tongue and jaw position. The figure shows that as movement amplitude for the final vowel decreases, there is an associated increase in the predicted spatial and temporal asynchrony of tongue and jaw motion.

normal sound level and rate. Fifteen samples of each utterance type were recorded. To date, three speakers have been tested.

4. RESULTS

A computer model of the tongue-jaw system with simplified geometry and dynamics was used to explore the patterns of coarticulation that might arise from a system of mechanically coupled masses under neuromuscular control. We used the same muscle model and simulated neural control as reported in Gribble et al. [2]. The geometrical configuration was highly simplifed. The jaw was connected to a fixed upper skull and fixed sternum. The tongue was connected to the upper skull and the jaw. As in our previous work, it was assumed that the movements of the tongue and jaw were produced by changes to neurally specified tongue and jaw equilibrium positions.

Figure 3 shows the simulated tongue and jaw motion kinematics (curved lines) and underlying control signals (dotted lines). In the two top panels, we show a simulated *VCV* sequence in which synchronized control signals to the tongue and the jaw produce closing then opening movements. Simulated vertical positions are shown.

In order to simulate anticipatory coarticulation, we have varied the rate and duration of the equilibrium shifts associated with the tongue and jaw opening movements *CV* transition) while holding constant the rate and duration of the equilibrium shifts for the initial closing movement. This is equivalent to varying the identity of the final vowel while holding the initial vowel constant. By examining the simulated kinematics under these conditions we can assess the extent to which kinematic patterns characteristic of anticipatory coarticulation may arise, when at the level of the central control signals there is no account taken of upcoming context.

In the top two panels of Figure 3, if one were to measure the simulated kinematic amplitude and duration of the initial closing phase (from zero velocity during the initial vowel to zero velocity during consonant closure), it would be seen that the amplitude and duration of the initial closing movement increases as the amplitude of the final opening movement decreases. Moreover, even when the control signals to the tongue and jaw are synchronized, the simulated tongue movements reach maximum elevation later than the jaw.

A sample of the empirical results for tongue and jaw movements is given in Figure 4. The figure shows data for one subject under conditions in which the initial vowel and consonant are fixed while the final vowel in varied (mean ± one standard error are shown).

The empirical patterns of coarticulation observed here are similar to the patterns predicted in the simulations. As in

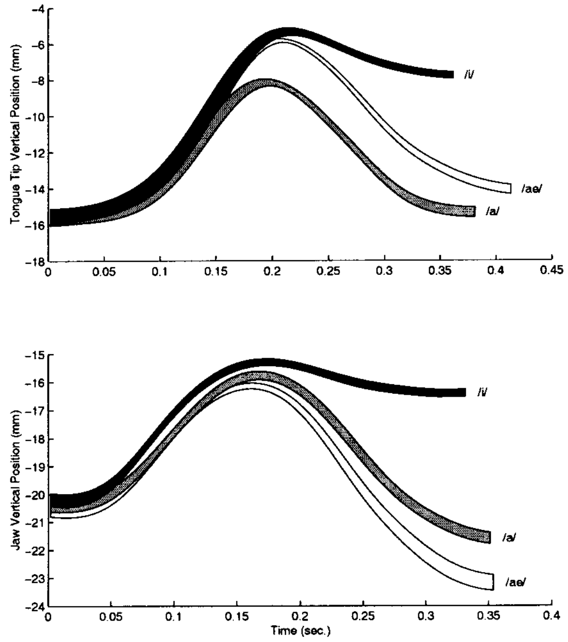

Figure 4: Empirical patterns of tongue and jaw movement match predictions of the model when there is no adjustment for context at the level of modeled control signals.

the simulations (in which the tongue and jaw closing control signals take no account of upcoming phonetic context), both the average amplitude and duration of the initial tongue and jaw closing movements vary inversely with the movement amplitude for the final vowel. Initial amplitudes and durations are greatest for the vowel *i* and least for the lower vowels. The timing of tongue and jaw movements also follows the pattern predicted in the simulations. The tongue reaches its maximum elevation later than the jaw, and the magnitude of the time difference is greatest when the following vowel is *i*.

5. DISCUSSION

This has been a preliminary examination of "anticipatory" coarticulation in tongue and jaw movement. We have shown in simulations that when no adjustment for changes in context occurs at the level of control signals, the predicted tongue and jaw motion kinematics differ as a function of context in a manner comparable to that observed empirically in anticipatory inter-articulator coarticulation.

In the modeling studies, these unplanned kinematic effects arise due to muscle properties and dynamics. However, our conclusion is not that coarticulation is unplanned. Rather care must be taken in drawing conclusions about the central planning processes underlying coarticulation without explicitly accounting for muscle properties and dynamics which may give rise some of these effects.

ACKNOWLEDGEMENTS

This research was supported by NIH grant DC-00594 from the National Institute on Deafness and Other Communication Disorders, NSERC-Canada, and FCAR-Québec.

References

[1] A. G. Feldman, S. V. Adamovich, D. J. Ostry, and J. R. Flanagan. The origin of electromyograms – Explanations based on the equilibrium point hypothesis. In J. Winters and S. Woo, editors, *Multiple Muscle Systems: Biomechanics and Movement Organization*, pages 195–213. Springer-Verlag, New York, 1990.

[2] P. L. Gribble, D. J. Ostry, V. Sanguineti, and R. Laboissière. Are complex control signals required for human arm movement. *Journal of Neurophysiology*, 79:1409–1424, 1998.

[3] R. Laboissière, D. J. Ostry, and A. G. Feldman. Control of multi-muscle systems: Human jaw and hyoid movements. *Biological Cybernetics*, 74:373–384, 1996.

[4] D. J. Ostry, P. L. Gribble, and V. L. Gracco. Coarticulation of jaw movements in speech production: Is context-sensitivity in speech kinematics centrally planned? *The Journal of Neuroscience*, 16:1570–1579, 1996.

[5] J. S. Perkell, M. Cohen, M. Svirsky, M. Matthies, I. Garabieta, and M. Jackson. Electromagnetic midsagittal articulometer (emma) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92:3078–3096, 1992.

[6] V. Sanguineti, R. Laboissière, and D. J. Ostry. A dynamic biomechanical model for neural control of speech production. *Journal of the Acoustical Society of America*, 103:1615–1627, 1998.

[7] F.E. Zajac. Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control. *CRC Critical Reviews in Biomedical Engineering*, 17:359–415, 1989.