1025

IS COARTICULATION IN SPEECH KINEMATICS CENTRALLY PLANNED?

David J. Ostry, Paul L. Gribble, and Vincent L. Gracco
McGill University, Canada and Haskins Laboratories, USA

Abstract

Coarticulation in speech production is a phenomenon in which the articulator movements for a given speech sound vary systematically with the surrounding sounds and their associated movements. Although these variations may appear to be centrally planned, 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. In the present paper, we address the origins of coarticulation by comparing the results of empirical and modeling studies of jaw motion in speech.

Resumé

Le phénomène de la coarticulation de la parole démontre que pour un son donné, les mouvements des articulateurs varient systématique avec les sons environnants et leur mouvements associés. Malgré que ces variations semblent être planifiées centralement, sans modèles explicites des articulateurs de la parole, les schémas cinématiques attribués à un contrôle central ne peuvent être distingués de ceux qui relèvent de la dynamique du système. Le présent article traite des origines de la coarticulation den comparant les résultats d'études empirique et de simulations du mouvement de la mâchoire dans le langage parlé.

Introduction

Speech production is a sensorimotor process in which neural representations of language are transformed into vocal tract motion. The sounds of speech may be combined in a variety of ways and the associated articulator movements may vary as the kinematic context changes. This kinematic variation, known as coarticulation, is one of the most pervasive characteristics of speech production. Some aspects of coarticulation may be centrally planned, while others may not be planned but may arise from factors such as muscle mechanics, musculo-skeletal geometry and dynamics. However, without explicit models of the speech articulators, kinematic patterns which are attributable to central control cannot be distinguished from those which arise due to muscle properties and dynamics and are not represented in the central control signals.

In the present paper we explore the possible origins of kinematic patterns of context sensitivity by comparing the results of empirical and modeling studies of human jaw motion. The simulations show that even when no account of context is taken at the level of central control, the kinematics may vary as a function of the preceding or following movement. The main point we wish to make is that unplanned effects may arise due to articulator mechanical and dynamic factors and these must be accounted for before drawing conclusions about the role of central control in coarticulation.

Jaw Model

We have recently proposed a model of jaw and hyoid motion based on the equilibrium point (EP) hypothesis of motor control (Laboissière, Ostry, and Feldman in press). The jaw model includes seven modeled muscle groups and four kinematic degrees of freedom: sagittal plane jaw rotation, horizontal jaw translation, hyoid vertical translation and hyoid horizontal translation. Consistent with empirical data, neural control signals to individual muscles (λs) are coordinated to achieve independent changes in the values of the system's four degrees of freedom (Ostry and Munhall 1994). Thus, control is organized to produce movements such as jaw rotation or translation, either alone or in combination.

In order to assess the extent to which muscle properties and dynamics contribute to the observed coarticulatory patterns, we have held the hypothetical central control signals fixed so that they take no account of context. Thus at a control level there is no planned coarticulation. We then examine the simulated kinematic patterns to assess whether the empirically observed patterns of coarticulation are nonetheless obtained. This in effect gives us a measure of the contribution of mechanics and dynamics to kinematic coarticulation.

Methods

Jaw motion kinematics were recorded during the repetitive production of VCV sequences in which the movement amplitudes for the initial and final vowels were varied. The movement amplitudes and durations of the jaw closing movement associated with the transition between the initial vowel and consonant were assessed as a function of the final vowel (anticipatory coarticulation). Amplitudes and durations of the opening movement from the consonant to the second vowel were assessed with respect to the initial vowel (carryover coarticulation).

The utterances were composed of the vowels *a*, *o* and *i* and consonants *k* and *t*. The utterances were embedded between flanking consonants *p* and *s* to produce speech-like sequences such as *sakas*. Jaw motions were recorded in 3D using Optotrak.

Results

Anticipatory Coarticulation

Figure 1 provides a summary of the main findings for anticipatory coarticulation using data for pitch as an example. It shows anticipatory coarticulation measured empirically and comparable patterns predicted by the model. The upper panels give the average empirical data for a single subject and comparable simulation results. The bottom panels give the empirically observed average movement amplitude and duration during the initial jaw closing phase (VC transition) as a function of the final vowel (for the data at the top left). For visualization purposes the functions in the upper left panel were time normalized before averaging and aligned for initial pitch angle.

Note that as predicted by the simulation (in which the jaw closing control signal takes no account of upcoming phonetic context), both the average amplitude and average duration of the initial jaw closing movement vary inversely with the movement amplitude associated with the final vowel. Initial amplitudes and durations are greatest for the vowel *i* and least for the vowel *a*.

Carryover Coarticulation

Figure 2 provides an example of carryover coarticulation. The upper panels give the average empirical data for a single subject and comparable simulation results, again using pitch as an example. The bottom two panels give the empirically observed average movement amplitude and duration during the jaw opening as a function of the initial vowel (for the data at the top left). As in Figure 1, for purposes of visualization, the

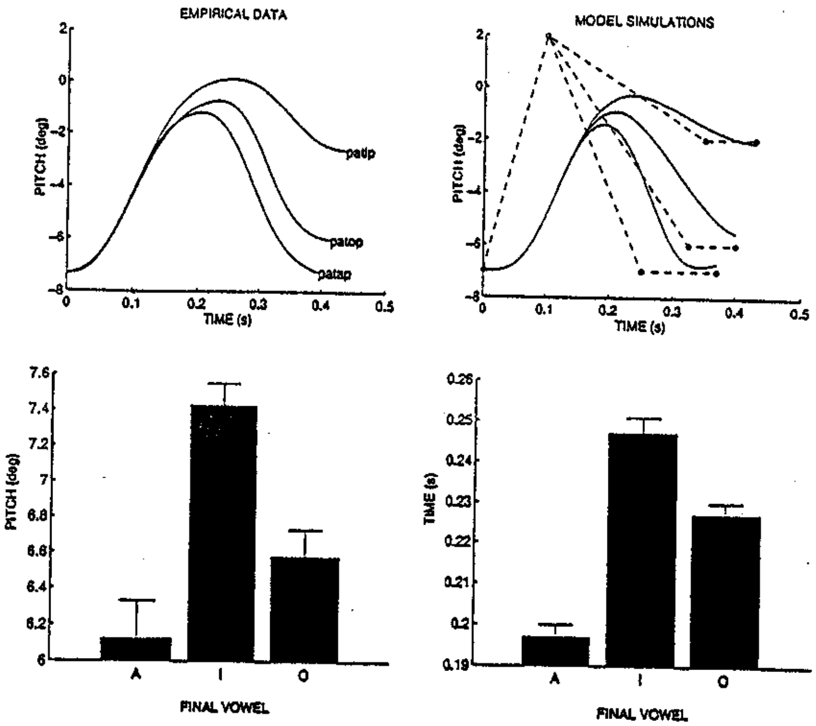

Figure 1: Comparison of empirical data and model predictions for anticipatory coarticulation.

functions in the top left panel were time normalized before averaging and aligned for final pitch angle.

Discussion

We have examined sources of "anticipatory" and "carryover" coarticulation in jaw motion by comparing the kinematics of vowel-consonant-vowel sequences with the results of modeling studies. In the simulations, we have shown that even when no adjustment for changes in context occurs at the level of central control signals, the predicted jaw motion kinematics differ as a function of context in a manner comparable to that observed in intra-articulator coarticulation. In the modeling studies, these unplanned kinematic effects arise due to muscle properties and dynamics. Accordingly, one should not draw conclusions about the central planning processes underlying coarticulation without explicitly accounting for these factors.

In the simulations, coarticulation arises as a consequence of the forces developed due to equilibrium shifts. As shown in Figure 1, the centrally specified equilibrium shift which gives rise to the movement occurs well in advance of the accompanying kinematic changes. The initial equilibrium shift towards the consonant position is followed by a subsequent shift back towards the equilibrium configuration for the final vowel. The second shift occurs while the articulator is still in the course of the first movement. The forces and torques which first close the jaw develop in proportion to the difference between the equilibrium and actual jaw positions. When the equilibrium shift begins back towards the final vowel, forces develop in the opposite direction and oppose the initial jaw closing forces. The magnitude of the opposing forces varies in proportion to the difference between the current equilibrium and actual positions. Thus as the rate of shift away from the equilibrium configuration for the consonant increases, for example, for *a* versus *i*, the magnitude of the opposing forces

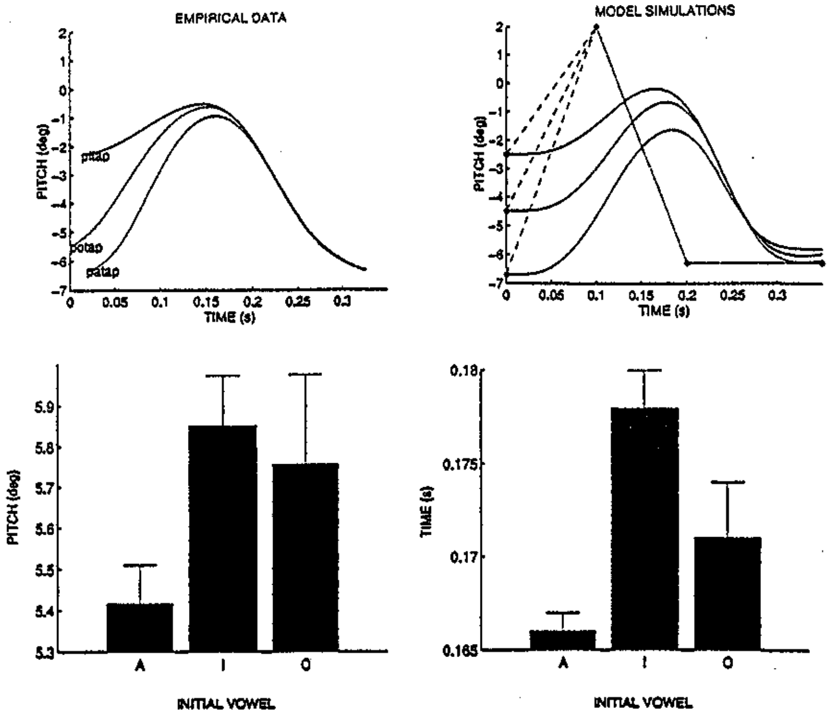

Figure 2: Comparison of empirical data and model predictions for carryover coarticulation.

also increases. The resulting effect is to reduce the net closing force more for large amplitude movements than for small movements. This leads to a greater reduction of both the amplitude and the duration of the initial movement towards the consonant for lower final vowels and thus accounts for the variation observed in the simulation results.

The sources of coarticulation revealed in the present study should be distinguished from previous accounts of coarticulation. Since the empirical variations observed may be unplanned, this demonstration should not be equated with the results of so-called "scan ahead" mechanisms (Henke 1966). Nor is the variation observed here equivalent to that in kinematic blending schemes in which it is proposed that coarticulation arises as the result of overlapping control signals for vowel and consonant related movements (Fowler 1977). As the present simulations show, this result is presumably due neither to a mixing of commands nor to an adjustment for context, but rather, as suggested above, to sequential control signals, muscle properties and dynamics.

References

Fowler, C. A. (1977). *Timing control in speech production.* Ph. D. thesis, Dartmouth College.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation.* Ph. D. thesis, M.I.T.

Laboissière, R., D. J. Ostry, and A. G. Feldman (in press). Control of multimuscle systems: Human jaw and hyoid movements. *Biological Cybernetics.*

Ostry, D. J. and K. G. Munhall (1994). Control of jaw orientation and position in mastication and speech. *Journal of Neurophysiology 71*, 1515–1532.