ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication

Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb26. Coarticulation in a whole event model of speech production

Bryan Gick*, Ian Stavness and Chenhao Chiu

*Corresponding author's address: Department of Linguistics, UBC, 2613 West Mall, Vancouver, V6T 1Z4, British Columbia, Canada, gick@mail.ubc.ca

Previous models of coarticulation have used varying combinations of advance planning and on-line calculation of weighted averages to determine how temporally overlapping speech sounds interact. A robust model of coarticulation should be able to predict such local interactions, as well as to describe changes resulting from degree of temporal overlap between adjacent events. Using a model of tongue-jaw-hyoid biomechanics (www.artisynth.org), the present paper demonstrates that typical cases of lingual coarticulation can be attributed to the intrinsic biomechanics of the human body in an entirely feed-forward model with no additional machinery. Biomechanical modeling outcomes are compared to speech production results from a previous articulometry study, and show that naturalistic coarticulatory patterns for VCV sequences emerge simply by temporally overlapping canonical muscle activations in a biomechanically realistic model. The built-in mechanics of the human body can handle at least simple VCV coarticulatory interactions with no extrinsic model at all, save one that identifies a) the right body parts, and b) the time-course of events.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Models of coarticulation typically handle local interactions between temporally overlapping consonants and vowels either through advance planning or algorithmic calculation of weighted averages, or some combination of these (see Farnetani and Recasens, 2010). Here we propose a third option, where the intrinsic biomechanics of the human body resolve local interactions in a feed-forward model.

Gick et al. (2012) describe a model in which muscular responses to articulatory and aerodynamic aspects of speech production interact. In this model, facial muscle actions associated with achieving articulatory goals (e.g., closing the lips) activate simultaneously with those responsible for containing intraoral air pressure (e.g., stiffening lips and surrounding perioral structures). These simultaneous activations are resolved via the body’s natural biomechanics, with no additional calculation or planning, as demonstrated in a realistic 3-D biomechanical simulation. We suggest that this same mechanism should apply to any local interaction between muscle activations. Coarticulation is one such case.

We use biomechanical modeling of the tongue and jaw to test whether temporally overlapping tongue and jaw muscle activations related to adjacent vowels and consonants are sufficient to simulate realistic coarticulatory interactions in VCV sequences. Simulation results were compared with electromagnetic articulometry (EMA) results reported by Recasens and Espinosa (2009).

SIMULATION

Simulations were carried out using a coupled tongue-jaw-hyoid model (see Stavness, Lloyd, Payan, and Fels 2011; Stavness, Lloyd, and Fels, 2012) in the ArtiSynth biomechanical simulation toolkit (www.artisynth.org). This model has been effectively used in previous studies of interactions between /r/ and vowels in English (Stavness, Gick, Derrick, and Fels, 2012). Fixed boundary conditions were used to connect the finite element model (FEM) tongue to muscle attachment points on the mandible and hyoid bone, and hydrostatic effects were simulated using constraint-based tissue incompressibility.

For the present study, we simulated 100% overlap between adjacent sounds, such that muscle activations for vowels in VCV sequences were maintained throughout the sequence. If complete overlap elicits realistic coarticulatory effects, then reducing the degree of overlap will quantitatively lessen coarticulation. Muscle activations from Stavness, Gick, Derrick, and Fels (2012) were used to simulate vowels [a] and [i]; activations for consonants [n] and [k] were determined heuristically and set to approximate known shapes for consonants in isolation. Thus, simulated sequences were [ana], [ini], [aka], and [iki]. Note that a key aspect of simulating 3-dimensional consonant postures is lateral bracing; it was a priority in determining activations for the present study that bracing be maintained for all postures except for [a]. Figure 1 shows a schematic view of combined activations for vowels (continuing throughout the sequence) and consonants (transitioning in the middle of the sequence).

FIGURE 1. Schematic of overlapping muscle activations for VCV sequences.
To simulate the EMA results reported by Recasens and Espinosa (2009), we tracked fleshpoints on the simulated tongue surface. Figure 2 shows simulated tongue/jaw/hyoid postures for VCV sequences.

**FIGURE 2.** Simulated postures from VCV sequences. Images (a) and (b) show postures for vowels [a] and [i], respectively; (c) and (d) show mid-consonant posture for [ata] and [iti]; (e) and (f) show postures for [aka] and [iki]. The red pellets visible on the tongue surface represent simulated EMA transducer coils used to track flesh point locations.

Simulations show that the tongue was indeed able to achieve mid-consonant closure in all VCV sequences, despite constant activation for vowel postures.

Figure 3 shows relative positions of the simulated EMA coils for both consonants in [i] vs. [a] contexts, as compared with actual EMA results for two Catalan speakers reported by Recasens and Espinosa (2009). Simulated coil locations closely follow the expected trend in tongue fleshpoint positions, with the [t] posture tilting up and back in the [a] context (relative to the [i] context), and with all [k] coils retracting in the [a] context (relative to the [i] context). Note that it would be possible to use data from coil locations such as these as input to an inverse model of muscle activation (following Stavness, Lloyd, and Fels, 2012). However, as the purpose of the present study was to observe the additive effect of overlapping idealized postural activations (rather than to match perfectly the results of any particular previous study), we did not use this method.

**DISCUSSION**

Our results indicate that plausible patterns of coarticulation may be obtained simply by simultaneously activating the muscles associated with temporally overlapping canonical consonants and vowels in a biomechanically realistic simulation environment. A strong interpretation of these results suggests that no model at all may be needed to describe coarticulatory processes (at least in the limited context considered in the present study) save a simple sliding timescale to specify overlap, and a realistic model of speech biomechanics.
A complete model of coarticulation should be able not only to predict such local interactions, but also to describe specific variation such as articulator-specific coarticulatory resistance/aggressiveness (Recasens and Espinosa 2009), as well as changes resulting from varying factors such as rate, which should influence (minimally) degree of temporal overlap between adjacent events. These effects are straightforward to simulate in our model, and will be the subject of future study.

The present study has provided suggestive evidence that the built-in mechanics of the human body can go a long way towards handling local coarticulatory interactions using no advance planning, contextual information, or extrinsic model, save that which provides knowledge of muscle activations for canonical postures and the temporal overlap of speech events.

ACKNOWLEDGMENTS

The authors gratefully acknowledge much helpful input from colleagues, including S. S. Fels, J. E. Lloyd, and D. H. Whalen. Research funded by an NSERC Discovery Grant to the first author.

REFERENCES


