Prosodic effects on articulatory movements at phrase boundaries in spontaneous speech

CASE
COLLEGE OF ARTS AND SCIENCES

Patrizia Bonaventura

Department of Communication Sciences
Case Western Reserve University

ABSTRACT

The present study aims to investigate whether phrase-final lengthening affects 'icebergs', defined by the C/D model (Fujimura, 2003) as stable patterns that have been shown to maintain their robustness in different prosodic environments (Fujimura, 1996). Previous studies (Bonaventura, 2003) have shown evidence in favor of the existence of 'iceberg' patterns, but a linear dependence of slope on the total excursion of the demisyllabic movement appeared, instead of the predicted constancy of speed at iceberg threshold crossing. Systematic outliers from the linear dependence of slope on excursion, suggested that the presence of a following long boundary might have induced "phrase-final elongation". This lengthening effect is manifested by an increased acoustic syllable duration and a decreased jaw opening, so that the iceberg crossing speed of the crucial articulator in the final demisyllable may be consequently slowed down. Boundary and excursion effects are important because they identify the inherent patterns characteristics of each demisyllabic phonetic segment (e.g. articulatory excursion is generally more limited in high and mid vowels; Erickson, 2002).

Previous studies on read data (Bonaventura and Fujimura, 2004) from a corpus of simulated dialogues (Erickson, et al. 1998) have shown no significant effect of syllable magnitude, measured as syllable duration and of boundary strength, measured as gap duration between syllables, on iceberg speed, probably due to the read nature of the dialogues, not allowing for relevant variations in excursion at phrase boundary.

The present study investigates the nature of the "phrase final elongation effect" by observing possible influence of syllable duration (as a measure of syllable magnitude), and of articulatory gap duration (as a measure of boundary strength) on the preceding movement pattern, focusing on microbeam data from a corpus of semi-spontaneous speech ('Blue Pine'; Menezes, 2003). The aim of this study is to verify whether the elongation effect can be identified and isolated in utterances presenting a larger variety of prosodic conditions because occurring in a more natural conversational context (Menezes, 2003; Mitchell et al., 2000).

Results show a prominent effect of excursion on speed, with no significant changes in speed due to magnitude of syllables or boundaries, . These results seem to show that prosodic effects due to boundary or syllable magnitude do not systematically affect speed of consonantal movements, and agree with previous findings that showed no role of syllable magnitude and boundary duration in the prediction of speed in read speech.

INTRODUCTION

The Converter/Distributor (C/D) Model, as a non-traditional, more powerful model of phonetic organization, uses syllables instead of phonemes as the concatenative units of speech signals. It represents the rhythmic organization of an utterance by a magnitude-controlled syllable-boundary pulse train. Based on the magnitude distribution of the syllable pulses, syllable durations are computed. Assuming temporal gaps between consecutive syllables to represent the boundary pulse magnitudes, syllable-boundary triangles are placed in time as contiguous series. The phonetic magnitude of each boundary as well as syllable is continuously variable according to various utterance factors.

ICEBERG PATTERNS

Fujimura [1986], based on microbeam data, suggested that a certain part of articulatory movement patterns was characteristically constant for a given demisyllable, when stress conditions varied. The speed of movement of the crucial articulator (responsible for an obstruent gesture) crossing a fixed height threshold (relative to the occlusal plane, for a given speaker) was identified as such a relatively invariant pattern. The time function representing the first time derivative (velocity) of a flesh-point (pellet) position resembled the tip of the floating iceberg.
Later, Fujimura proposed the C/D model as a new comprehensive theory of phonetic implementation. It assumed (like Öhman [1967]) that a sequence of vowel gestures for syllable nuclei formed a slowly changing syllabic gesture (phonetic status contour) as an aspect of what is called base function, on which local quick gestures for consonants (elemental gestures) are superimposed, according to syllabic feature specifications.
The C/D model further assumes that each elemental consonantal gesture constitutes a fixed ballistic motion pattern, which, as a passive response, is evoked by a time-shifted replica (onset pulse, coda pulse, etc.) of the syllable pulse as the excitation.
For computation of the phonetic implementation, all elemental gestures for consonants as demisyllabic constituents (*i. e.*, onset, coda, or syllable affix) of the syllable, are stored in an impulse response function (IRF) table.

PHRASE FINAL ELONGATION

Phrase-final elongation is a phonetic phrase boundary effects, consisting in rhyme elongation in phrase-final position. The elongation could be modeled as an expansion of the duration of the gestures, slowing down all gestures in the last syllable. It could result in inserting specific boundary duration, delaying the occurences of coda and s-fix gestures, or it could create a pause, whether it is a period of silence or a period filled with spilled over voicing and other articulatory gestures. It could be a combination of these. Currently available articulatory data are not sufficient to determine which model should be used. Some deviations from the linear relationship that characterizes the iceberg slopes, has been found, especially in digits at phrase boundaries (Fig. 3, Bonaventura, 2003). Previous studies on read speech have found that no effect of syllable duration and gap duration affect the speed of movement at threshold (Bonaventura and Fujimura, 2004). The present study aims to investigate and quantify the effect of syllable magnitude and boundary strength on speed of the consonantal movements in the selected syllables, in a spontaneous speech corpus, where presence of phrase boundary effect has been identified by previous research (Menezes, 2003).

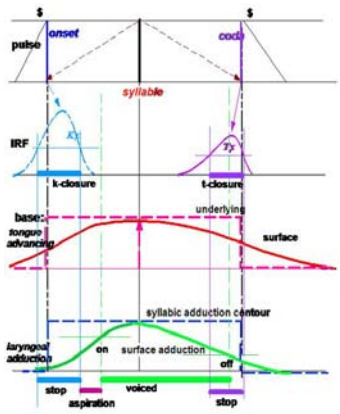


Figure 1: 'kit' (C/D diagram)
(courtesy of Prof. O. Fujimura)

METHOD

Data and subjects

For the present study, the 'Blue Pine' corpus was used, collected by Donna Erickson (Erickson et al., 1998). The articulatory data were acquired at the University of Wisconsin by the X-ray Microbeam system. The data consist of articulatory records, obtained by automatically tracking the movement of gold pellets affixed on selected points of articulatory organs. Pellets were affixed on 4 points of the tongue surface, the lower lip, and on the mandible at the lower incisor and a molar tooth, recording the movement of the articulators with respect to the maxillary occlusal plane (bite plane). Head movements are corrected for, so that articulatory movements relative to the skull were measured.

The 'Blue Pine' data acquisition protocol was similar to Labov's paradigm (1972): the subjects were involved in typical dialogues consisting of a series of exchanges started by the experimenter, asking questions like: 'Where do you live?' and 'Where do you work?' The answer was of the type: the form "I work at 995 Pine Street. The correct street address was indicated to the subjects on a monitor but subjects were told not to read from the monitor, but to interact naturally with the experimenter. Subjects were also told that the experimenter was in a noisy environment and had to ask for clarification. During the recording, the experimenter sat behind the subject and out of sight (in the same room). A dialogue consisted of a series of exchanges between the experimenter and the subject, of the form: "Do you work at 559 Pine Street? No, I work at 955 Pine Street. Do you work at 995 Pine Street? No, I work at 595 Pine Street". This paradigm was used to investigate the effect of contrastive emphasis, along with other paralinguistic phenomena, on the implementation of articulatory gestures.

The corrected digit in the 3-digit sequence sounded emphasized, whereas the uncorrected digits, in the correcting utterances, were not emphasized. Subjects were two native speakers of Midwest American English, one females and one male, who had no awareness of the purpose of the experiment.

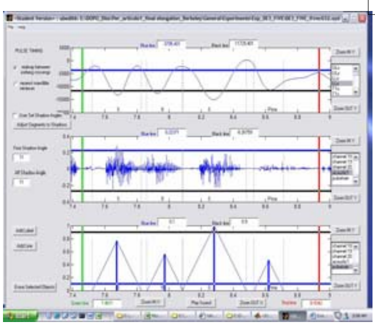


Figure 2 . Program 'ubedit': segmentation on the Lower Lip y channel (upper window), waveform and syllable triangles. Last phrase "5 9 5 Pine"

Data analysis

Prosodic effects on iceberg invariance (i.e. on speed at position threshold), were measured as duration of the syllable and duration of the articulatory gap following the digit.
Segmentation and measurements of digits was performed by the program 'ubedit', designed and implemented by Bryan Pardo; the example in Fig. 2 shows segmentation on the Lower Lip vertical movement tracking in the upper window, the corresponding waveform and syllable triangles in the lower two windows. The phrase analyzed is "5 9 5 Pine" (from speaker DE1, dialogue 6).
The data observed represent the speed of movement of a flesh-point on the tongue blade or the lower lip as the crucial articulator of the observed consonant, when the flesh point crosses a fixed vertical position (iceberg threshold) relative to the occlusal plane.

28 productions of 'five' and 32 productions of 'nine' were observed, realized by two speakers.

RESULTS

A multiple regression was conducted to evaluate how well excursion, syllable duration and gap duration, would predict speed of consonantal movements in initial and final demisyllables. The linear combination of the predictors was significantly correlated to the speed index, F(3, 24) = 5.84, p = .004 for initial demisiyllable in 'five'; F(3, 24) = 11.23, p = .001 for final demisyllable in 'five'; F(3, 25) = 5.878, p < .01 for intial demisyllable in 'nine' and F(3, 28) = 8.46, p < .001 for final demisyllable in 'nine'. The sample multiple correlation coefficients were .42 for 'five' initial demisyllable, indicating that 18% of the variance of the speed index in the sample can be accounted for by the linear combination of excursion, syllable duration and gap duration; .56 for 'five' final demisyllable (covering 31% of the variance), .41 for 'nine' initial demisyllable (covering 17% of the variance) and .48 for 'nine' final demisyllable (covering 12% of the variance).

All speakers 'five' initial dm	Bivariate	Partial
Excursion Syll. dur. Gap dur.	-.6* -.55* -.34***	-- .421*** -.2 -.081

*p< .001 for bivariate, *** p < . 05 for bivariate, ***p < .05 , both bivariate and partial

Table 1: Correlation coefficients from bivariate and multiple regressions between excursion, syllable duration and gap duration with speed for 'five', initial demisyllable

All speakers 'five' initial dm	Bivariate	Partial
Excursion Syllable dur. Gap duration	.707* .156 -.007	.753* -.336 -.022

*p < .001

Table 2: Correlation coefficients from bivariate and multiple regressions between excursion, syllable duration and gap duration with speed for 'five', final demisyllable

All speakers 'five' initial dm	Bivariate	Partial
Excursion Syllable dur. Gap duration	-.58* -.29 .005	-.533** -.29 -.338

*p < .001 for bivariate, p < .001 for partial

Table 3: Correlation coefficients from bivariate and multiple regressions between excursion, syllable duration and gap duration with speed for 'nine', initial demisyllable

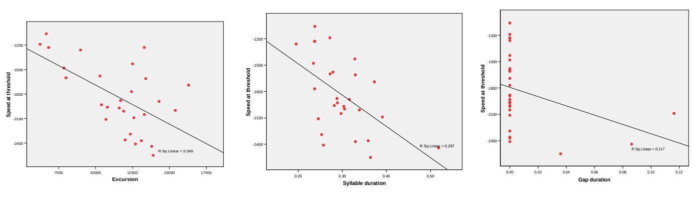
all speakers 'five' initial dm	Bivariate	Partial
Excursion Syllable dur. Gap duration	.63* -.117 -.13	.683* -.221 -.06

*p < .001 for bivariate, p < .001 for partial

Table 4: Correlation coefficients from bivariate and multiple regressions between excursion, syllable duration and gap duration with speed for 'nine', final demisyllable

All speakers - 'five'– initial dm	Z predicted speed = -.44 z exc + -.25 z syll dur + -.09 z art gap
All speakers - 'five'– final dm	Z predicted speed = .909 z exc + -.368 z syll dur + .019 z art gap
All speakers - 'nine'– initial dm	Z predicted speed = -.532 z exc + -.354 z syll dur + .394 z art gap
All speakers - 'nine'– final dm	Z predicted speed = .7 z exc + - .25 z syll dur + - .6 z art gap

Table 5: **Relative importance of predictors for speed at iceberg threshold;** standardized regression coefficients and prediction equations for the standardized variables: excursion, syllable duration, gap duration



Tables 6-7-8: **Derivative at threshold vs. excursion, articulatory syllable duration, and gap duration, all speakers, 'five', initial demisyllable**

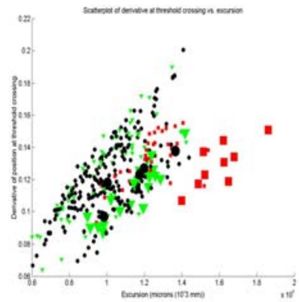


Figure 3: Derivative at crossing vs. excursion, speaker DE3, lower lip vertical displacement, final demisyllable, word 'five' (340 curves), read speech

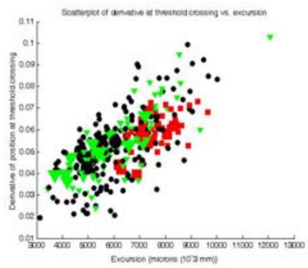


Figure 4: Speaker DE3 , scatterplot showing maximum speed vs. excursion for 340 iceberg curves, tongue blade vertical displacement, final demisyllable, word 'five' , read speech

CONCLUSIONS

The outcome of the correlation analyses seem to confirm previous results from read data, showing a prominent role of excursion in prediction of iceberg speed in lower lip and tongue tip consonantal movements. Syllable duration and gap duration do not show in this spontaneous speech corpus, a significant role as predictors of speed, therefore, no systematic influence of syllable magnitude and boundary duration appears in read or spontaneous speech.

Also, deviation from the linear dependence of movement speed on excursion, appearing in final demisyllables for the lower lip in 'five' in read speech (Bonaventura, 2003), has not been found for the tongue blade, in same context (Fig. 4, Bonaventura, 2006). Rather than being a systematic deviation due to phrase final elongation, maybe the reduced variation in speed in phrase final position, might be considered as an individual strategy adopted locally. Therefore, it seems that subjective factors in the realization of muscle trajectories in syllable final position before phrase boundary, have to be taken into account in modeling iceberg patterns.