Stable patterns of articulatory movements across inter-subject variability

Patrizia Bonaventura* and Paul Taylor**

**Department of Communication Sciences*
Case Western Reserve University
***Quality Electrodynamics*

CASE
COLLEGE OF ARTS AND SCIENCES

ABSTRACT

The goal of the present study was to test whether models of portions of curves, representing movements of the crucial articulator for production of place in syllables containing labiodental and alveolar gestures for production of obstruents ('iceberg' portions of demisyllables), that had previously been found to be stable across different prosodic conditions (Bonaventura, 2003; 2005; 2006; Bonaventura and Fujimura, 2007), a) remained stable across different subjects pronunciations, for each consonantal class b) were significantly different for the two different consonantal movements. Curves were previously extracted from microbeam articulatory data, from 3 subjects for Lower Lip movement (LL) for word 'five' and 3 subjects for Tongue Tip (TT) movement for word 'nine'.

Curve fitting models were obtained, by using a best fit fourth order polynomial, from a total of 1193 curves representing lower lip vertical displacement for production of [f] and [v] in 'five' and from a total of 610 curves representing tongue tip vertical displacement for production of [n] in 'nine'. Coefficients were statistically compared, to verify presence of a) non significant difference between models across pronunciations by 3 subjects b) significant difference between the two generalized curves for [f] vs. [t] across subjects.

Positive results from (a) would support the hypothesis of presence of articulatory pattern that would remain stable across different prosodic conditions and inter-subject variability, possibly indicating properties of an identifiable articulatory unit. Positive results from (b) would possibly indicate a consistent difference between crucial articulator movements for production of labiodental vs. dental consonantal gestures.

Results showed no expected similarity between 'movement curves across subjects pronunciations, except for some stability in the TT coefficients in coda. However, the comparison between the coefficients of the generalized models for TT and LL showed significant differences between the two movements in final demisyllable. These results partially confirm the expected difference between models, indicating that at least the fitted curves for TT and for LL in final demisyllable, if more stable across subjects realizations, could be considered as a reference pattern, representing normal speech for comparison with abnormal production.

INTRODUCTION

Consistent patterns of kinematic and physiological characteristics of speech movements have been long investigated, in order to identify strategies to compute articulatory trajectories and to reliably predict displacement and timing patterns of units of movement, that can be used as reference, for example, to compare normal with disordered speech patterns. A short review of previous studies will present, in the following, the most common approaches to analysis of kinematic parameters for characterization of displacement curves for selected articulators and of their timing properties.

Spatio-temporal parameters in movement units characterization: displacement functions

Previous studies (Kent and Moll, 1972a), based on phonemic units, attempted to provide "a point parametrization of the articulatory apparatus which could lead to (1) the determination of the velocities for various points on an articulator, (2) a description of the variability in the articulatory positions associated with a given phoneme or speech event, and (3) an assessment of the distributed physiological mechanical constraints on an articulatory surface". Kent and Moll (1972a) measured different parameters relative to displacement of jaw and tongue, for production of vowels and consonants in VCV sequences under varying phonetic contexts, in 3 sentences, as produced by 2 speakers, at two different speaking rates, in order to obtain a point parametrization of the vocal tract. Articulatory positions of the tongue point and tongue body were observed in 3 sentences, as produced by 2 speakers, at two different speaking rates, and were compared graphically, by plotting the three position points in time. "The results from this experiment relative to tongue movement, suggested that for the articulations of /j,d,z/, but not /g/, individual points on the tongue assume relatively invariant positions in the oral cavity for the production of these sounds" .

Tasko and McClean (2004) attempted to identify kinematic parameters that would stay stable across different speech tasks and that could "help define a kinematic performance space for a given speaker." Average parameters as peak speed, distance and duration were measured in portions of movements, called 'strokes', defined as the period bounded by two successive minima in the speed history. Thus, a stroke is operationally defined as a single period, an acceleration and a deceleratiion. Overall spatial variability of these parameters was evaluated for each different task (i.e. variables produced in a spontaneous monologue or dialogue, in a reading task, in elicited repetitions, or in speech vs. non-speech). Results showed correspondence of jaw and tongue point fleshpoint strokes with acoustic events representing alveolar fricatives, but no systematic association between particular acoustic events and movement strokes. Also, marked differences between amplitude, speed and excursion between speech and non-speech task were found, as well as changes in spatial variability related to articulator type and to task.

Variability of movement displacement functions with speaking rate

Attempts to characterize articulatory trajectories (and hence relative speech motor control characteristics), have not formulated so far generalizable, linguistically meaningful criteria to classify portions of movement units that could be systematically associated with a distinctive linguistic unit. Such variability in results also derives from the fact that the different assumptions adopted with regards to the target units of the speech implementation process, led to selection of different units of observation and to different interpretations of the results found.

In speech, unlike for other limb movements (Cooke, 1982; Feldman, 1982), variations in rate generate complex changes in the muscle kinematics (Kuehn and Moll, 1976): some studies have shown a reduction in articulator movement amplitude and no change in velocity, or even a reduction in velocity at the increase of speaking rate (Kent and Moll, 1972), some an increase in velocity (Gay, 1981), some found increases in peak velocity with amplitude unchanged (Abbs, 1973). Some studies have therefore, attempted to examine the velocity/amplitude relationship in speech movements under controlled variations in amplitude due to changes in speaking rate. Some generalizations emerged from such studies, which are reported in the following.

Spatio-temporal parameters in movement units characterization: velocity/amplitude relationship

Some stable characteristics were found in kinematic behavioral patterns similar between speech and non-speech movements: "in both, there appear to be systematic changes in the slope of the relationship between maximum velocity and movement amplitude, a kinematic indicator of articulator stiffness, with differences in movement rate and duration" (Ostry and Munhall, 1985).

However, whereas in limb movements an increase in rate results in the increase of the slope of the relationship between maximum speed and movement amplitude (Cooke, 1980), increasing speaking rate did not show to generate the same amount of increase of slope as in limb movement: Ohala et al. (1968) observed a variation of the peak velocity of the jaw during a closing gesture for consonants, proportional with the amount of preconsonantal law opening; however, Kent and Moll (1972a), found such a relationship to appear clearly mostly at moderate utterance rate.

The less stable increase of the slope of the relationship between maximum speed and movement amplitude in speech movements, at the increase of rate, might be possibly due to the durational changes induced by variations in rate, stress and in phonetic context. Ostry et al. (1983; 1984) showed dependence of slope of the maximum-velocity/movement amplitude regression, mainly on stress, and only slight dependence on rate. However, Ostry and Munhall (1985) showed that the ratio between extent and maximum velocity, constituting an indicator of stiffness, held through controlled prosodic variations, and showed to vary inversely with the duration of the movement.

ICEBERG PATTERNS

Other studies, carried out based on the assumption of articulatory gestures constituting the basic units of speech implementation (Fujimura and Williams, 1999; Fujimura, 2003), observed the effects of controlled rate changes on the slope of the relationship of amplitude and velocity at a point close, but not coinciding, with the point of maximum velocity, called 'iceberg threshold'. The 'iceberg threshold' was an optimally determined position (i.e. the slope of the time function for position) where the velocity resulted, from previous experiments, to be relatively invariant for different prosodic conditions, given the demisyllable (Fujimura and Spencer 1983; Fujimura 1986b, 1996b; Mitchell, 2000). Results from the original studies on 'icebergs' (Fujimura, 1981a, 1983, 1986b, 1996b) are exemplified in Fig. 1 (for identification criteria for the 'iceberg' threshold, see Fujimura, 1996 and Mitchell, 2000; for justification of robustness and significance of the 'iceberg' threshold as a kinematic parameter, with respect to the most commonly adopted peak velocity point, see Bonaventura, 2003; and Bonaventura, submitted).

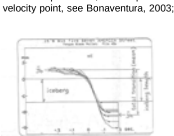

Fig. 1 Original 'iceberg' description (Fujimura, 1983)

Portions of gestures were studied ('iceberg' patterns) measured on X-ray microbeam data, and defined as the portion of monotonically increasing (VC) or decreasing (CV) pellet height of the crucial articulator of the obstruent gesture involved in each demisyllable of a syllable (in this study, /fajv/ and /najn/, corresponding to the monosyllabic words 'five' and 'nine'). In a first preliminary study (Fujimura, 1981a), observations of the movements as time functions, showed that the fastest parts of the movement patterns (icebergs') seemed to be invariant in slope (see Fig. 1) across changes in amplitude due to different prosodic conditions. As an interpretation of these results, it was hypothesized that, differently from the slope of the 'iceberg' curves, when measured at a time threshold, which was supposed to vary in proportion to the changes in amplitude of the syllable (determined by the amplitude of the jaw movement; Fujimura, 1996), the slope at a fixed position along the y-axis ('iceberg' threshold) did not vary for the same prosodic manipulation (" 'iceberg' invariance" hypothesis).

Since the highly nonlinear and complex articulatory system of mapping from muscle control to flesh-point position is difficult to depict by a simple mathematical model, it remained an empirical issue to determine whether velocity of 'iceberg' patterns remained constant at the fixed position 'iceberg' threshold under changes due to prosodic variations, and could, therefore, be used as a reference stable articulatory patterns to represent normal speech production.

GOAL OF THE STUDY

The examination of the kinematic properties of the demisyllabic movement patterns as time functions for flesh-point (pellet) position, for the same demisyllable under different prosodic conditions, was carried out on a larger corpus designed to contain prosodically varying speech material, in order to test of the validity of the 'iceberg' invariance hypothesis.

The position values corresponding to the monotonic part of the 'iceberg' curves were plotted, for all curves, by speaker (see an example relative to lower lip movement, for word 'five', in final demisyllable by speaker DE3, in Fig. 2).

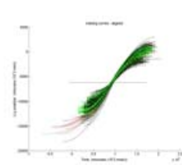

Fig. 2 Speaker DE3, word 'five' , 340 Iceberg curves – final demisyllable (Bonaventura, 2003)

Scatterplots were obtained, by plotting the first time derivative at threshold-crossing time against excursion of the curve (defined on the vertical distance traveled by the pellet from the beginning of each movement to its end as recorded; the end points of each curve were determined by detecting zero vertical velocity in the monotonic curve) (see an example relative to lower lip movement, for word 'five', in final demisyllable by speaker DE3, in Fig. 3).

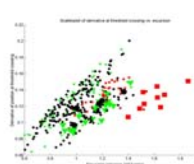

Fig. 3 Scatterplot showing derivative at threshold crossing vs. excursion: speaker DE3, word 'five' - final demisyllable (340 curves) (Bonaventura, 2003)

The results showed evidence in favor of the existence of 'iceberg' patterns, but a linear dependence of slope on the total excursion of the demisyllabic movement, instead of the approximate constancy of the threshold crossing speed as suggested in the original proposal of the 'iceberg', has been found, except for the longest curves, the ones referring to the realizations of the demisyllables in phrase final position, showing lengthening, hence greater amplitude and less strong linear relationship with speed.

Larger excursion in consonantal gestures has been observed in phrase-final and -initial, than in phrase-medial positions (Cho 2005, Bonaventura 2003, Keating et al. 2003, Kelso et al. 1985, Ostry and Munhall 1985). As shown in previous research (Fougeron and Keating, 1997), the magnitude and duration of consonantal gestures depend on syllable boundaries, position in the phrase, and strength of phrase boundaries. Further research has confirmed these results, showing larger displacements and longer relative durations of consonantal movements during constriction formation and release at phrase edges than in phrase-medial positions (Beckman and Edwards 1992, Byrd and Saltzman 1998, Byrd et al. 2000).

When boundary effects did not confound the movement speed variation pattern, a linear relationship was found between speed and excursion of tongue tip and lower lip movements; such systematic pattern, occurring at the 'iceberg' threshold, differed from previous results, obtained from observation of excursion variations at the maximum speed point, in which proportional change was found in the amplitude /velocity relationship.

The goal of the present study was to observe whether stable kinematic parameters, representing generalized position functions for the tongue tip and lower lip movements, could be identified around the 'iceberg' threshold, and if such recurrent parameters could be used to identify recurrent displacement patterns for selected articulators.

METHOD

Data and subjects

The articulatory data were recorded at the x-ray Microbeam Facility at the University of Wisconsin, Madison (Nadler et al. 1987; Fujimura et al., 1973; Kiritani et al., 1975), and include both articulatory and acoustic recordings: the former provide the tracings of movements of the speaker's articulators (tongue tip, body and dorsum, lower lip, and mandible, at a molar and the lower incisors). Head movements were corrected for, so that articulatory movements relative to the skull could be measured.

The data, the 'New Red Pine 96' corpus (Erickson et al., 1998), includes dialogues of the form "Is it 995 Pine Street? No, it's 955 Pine Street. Is it 995 Pine Street? No, it's 595 Pine Street." In particular, the questions were referred to as 'reference utterances'. Also, 'yes' answers are considered 'reference utterances'. The 'no' answers contained both emphasized digits (referred to as 'corrected'), and non-emphasized digits (referred to as 'uncorrected'). Both the reference utterances and the answers (yes or no utterances), were analyzed. The corrected digit in the 3-digit sequence was maximally emphasized, with respect to the uncorrected ones: the dialogues were designed to control for different prosodic conditions.

The digits corrected by the subject (either 'five' or 'nine'), were monosyllables, containing the low-vowel diphthong [aJ], selected to observe prosodic effects, independently from vowel effects: these data revealed in previous studies significant changes in jaw position, for emphasized vs. unemphasized (corrected) productions of the same digits (Westbury and Fujimura, 1989; Erickson and Fujimura, 1996; Menezes et al., 2002). The sentences were pronounced by three subjects, two males and a female, all speakers of Midwestern American English (Erickson and Fujimura, 1996).

Iceberg curves were measured and extracted from the tracings of the crucial articulator for production of the digits in the dialogues, 9 and 5 (see Fig. 4). These words are monosyllabic, they share the same obstruent articulator in initial and final demisyllable (Lower Lip for 'five' and Tongue Tip for 'nine'), they share the same vocalic nucleus (the diphthong /aj/). The number of curves observed, by word, speaker and demisyllable position, is reported in Table 1.

Speaker / # of curves	DE1	DE2	DE3
'five' - Initial dm	198	54	340
'five' - Final dm	207	55	339
'nine' - Initial dm	189	40	70
'nine' - Final dm	199	40	72
TOTAL	793	189	821

Table 1: Number of position curves modeled for the present study, by speaker, word and demisyllable position

Curve fitting models were obtained from a total of 592 curves representing Lower Lip (LL) vertical displacement for production of [f] in 'five' and from a total of 299 curves representing Tongue Tip (TT) vertical displacement for production of [n] in 'nine'. Best curve fitting resulted from fourth order polynomial models, providing a slightly better fit, especially around the 'iceberg' threshold, with the original 'iceberg' patterns. Fig. 4 shows results from second order (red), third (blue), fourth (cyan) and fifth (magenta) order polynomials models.

Selected fourth order coefficients, in the format of a b c d e (where y=a*t^4+b*t^3+c*t^2+d*t+e), for the Lower Lip and Tongue Tip 'iceberg' patterns model curves, by speaker (de1, de2 and de3) and demisyllable position (I vs. II, or onset vs. coda) , are reported in Table 2.

Fig. 4 Speaker DE3, word 'five' , 340 final demisyllable, curve fitting models of Iceberg curves (second order (red), third (blue), fourth (cyan) and fifth (magenta) order polynomials models.)

Tongue Tip movement	a	b	c	d	e
de1 I	2.12E-17	4.59E-12	-2.76E-07	-0.1093	9.18E+03
de2 I	8.16E-17	1.26E-11	-5.58E-07	-0.1515	7.14E+03
de3 I	9.53E-17	1.09E-11	-7.28E-07	-0.1503	4.72E+03
de1 II	-2.71E-17	-7.29E-12	-2.39E-08	0.1208	1.02E+04
de2 II	-9.35E-17	-1.73E-11	-2.96E-07	0.1515	9.36E+03
de3 II	-6.68E-17	-1.26E-11	-2.59E-07	0.1362	6.26E+03
Lower Lip movement					
de1 I	6.48E-18	2.26E-12	-2.03E-07	-0.0839	2.93E+03
de2 I	1.43E-17	5.64E-12	-3.26E-07	-0.113	8.11E+03
de3 I	-5.59E-18	2.99E-12	-2.22E-07	-0.0912	5.71E+03
de1 II	-4.38E-18	-2.18E-12	-1.21E-07	0.0759	2.01E+03
de2 II	1.42E-18	-2.31E-12	-6.29E-08	0.0991	8.24E+03
de3 II	-1.25E-17	-3.82E-12	-5.56E-08	0.1071	5.54E+03

Table 2 Fourth order coefficients for fitting models of Lower Lip and Tongue Tip 'iceberg' patterns

RESULTS

In order to test whether articulatory patterns could be found around the 'iceberg' threshold, that remained stable across different prosodic conditions and inter-subject variability, the selected coefficients were statistically compared, to verify presence of

a) non-significant difference between models of the Tongue Tip (TT) and of the Lower Lip (LL) across pronunciations by the 3 subjects

b) significant difference between the two generalized models for LL vs. TT movements, for each subject.

The former comparison was carried out on the basis of coefficients from all models of individual curves (see Table 1), whereas the latter analysis was carried out on coefficients from the selected models representing the TT and LL 'iceberg' patterns, respectively (see Table 2).

An independent samples t-test was conducted to evaluate the hypothesis that 'iceberg' curves showed no significant difference between coefficients for same articulator movement, across pronunciations with different prosodic patterns and by different speakers. Since the number of curves measured was different across subjects, the total number of data was reduced to 54 for LL movement and to 40 for Tongue Tip movement. The test was significant, except for the coefficients reported in Table 3**.**

LL - Idm	I coeff	II coeff	III coeff	IV coeff	V coeff
DE1-DE2	p = .001	p =.001	p =.001	p =.001	p = .001
DE2-DE3	p =.001	p =.001	**n.s.**	p =.001	p = .001
DE1-DE3	p =.001	p = .038	p =.004	p =.001	p = .001

LL - IIdm	I coeff	II coeff	III coeff	IV coeff	V coeff
DE1-DE2	**n.s.**	**n.s.**	p = .001	p = .001	p = .001
DE2-DE3	p = .001	p = .001	p = .001	p = .001	p = .001
DE1-DE3	p = .001	p = .001	n.s.	p = .001	p = .001

TT - Idm	I coeff	II coeff	III coeff	IV coeff	V coeff
DE1-DE2	p = .001	p = .001	p = .001	p = .001	p = .001
DE2-DE3	p = .001	p = .001	p = .001	p = .001	p = .001
DE1-DE3	p = .001	p = .001	**n.s.**	**n.s.**	**n.s.**

TT - IIdm	I coeff	II coeff	III coeff	IV coeff	V coeff
DE1-DE2	**n.s.**	**n.s.**	p = .001	p =.01	p = .001
DE2-DE3	**n.s.**	p = .001	p = .01	**n.s.**	p = .001
DE1-DE3	**n.s.**	p = .001	p = .001	**n.s.**	p = .001

Table 3: Differences between coefficients of curves, modeling original TT and LL movements position functions, contrasting speakers' realizations in different demisyllables

General patterns of similarities did not emerge from the t-test, even though most common coefficients among speakers, seemed to occur in the realizations of the TT movement in coda.

Independent t-tests were also conducted to evaluate the hypothesis that coefficients for generalized model curves for the TT movement, would differ significantly from coefficients for generalized model curves for the LL movement, across speakers, in onset (initial demisyllable) and in coda (final demisyllable). Significant difference (at least at .05 level, see Table 4) was found for all coefficients, except for the 3rd one, for the general model curves representing the TT and LL movements, produced by all subjects pronouncing same consonant, both in initial and final demisyllable.

TT vs. LL	I coeff	II coeff	III coeff	IV coeff	V coeff
I dm	**n.s.**	**n.s.**	**n.s.**	**n.s.**	p = .01
TT vs. LL	I coeff	II coeff	III coeff	IV coeff	V coeff
II dm	.05	.05	**n.s.**	.05	.01

Table 4: Differences among coefficients of generalized model curves for Tongue Tip vs. Lower Lip movements, for all speakers, in initial (I) vs. final (II) demisyllable

The overall dispersion of the curves around the calculated models (reported in Table 2), was evaluated by calculating the distance times time for every point of the models of the original position functions with respect to each relative model curve. Two examples are reported in Figg. 5 and 6.

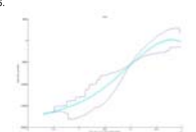

Fig. 5: Displacement area for 340 'iceberg' curves – final demisyllable, speaker DE3, word 'five', with respect to the model curve (middle cyan)

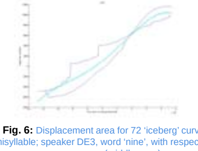

Fig. 6: Displacement area for 72 'iceberg' curves – final demisyllable; speaker DE3, word 'nine', with respect to the model curve (middle cyan)

The calculation was performed in order to verify whether common patterns of displacement would be found for same movements, across speakers. The values for the displacement, however, do not seem to show any common pattern across movements, speakers, or demisyllables, confirming the great inter-subject variability in realization of syllables in different emphasis and rate conditions.

CONCLUSIONS AND DISCUSSION

The results did not confirm the first hypothesis, showing no common pattern of displacement across speakers, for movements of Tongue Tip and Lower Lip, when measured around the 'iceberg' threshold; this effect was probably due to the fact that the movements have been realized with much variation, in the original corpus (interpretation confirmed by dispersion areas plots, Figg. 5 and 6), designed to generate pronunciations differing in prosodic characteristics. A further extension of this study would imply comparison of model coefficients for curves produced with same emphasis, or speaking rate, by several speakers, to verify whether position functions would result more consistent.

In the second analysis, the expected difference between the coefficients of the two models for the TT vs. LL displacement was found only for final demisyllable, except for the 3[rd] coefficient, representing the portion of the curve around the 'iceberg' threshold.

The results partially confirm the expected difference between models, indicating that at least the modeled curves (see Table 2) for TT and for LL in final demisyllable, if more stable across subjects realizations, could be considered as a reference patterns representing normal speech for comparison with abnormal production.

The results further suggest that the fastest part of each model, around the 'iceberg' threshold, show some stable behavior across different movements, speakers and prosodic realizations. Further analysis of modeled curves, possibly with an appropriate weighting window centering around the threshold crossing point, might provide an estimate of how, in the vicinity of the crossing point, the curve deviates from a straight line.