Keller, E. (1987). Factors underlying tongue articulation in speech. *Journal of Speech and Hearing Research, 30,* 223-229.

Factors underlying tongue articulation in speech

Eric Keller
Université du Québec à Montréal,
Centre Hospitalier Côte-des-Neiges

A factor analysis of 11 measures of lingual activity involved in the production of the syllable /ka/ showed that three factors can explain about 75% of the variance in the data. Measures of displacement and velocity loaded strongly on the first factor, durational measures loaded strongly on the second factor, and mid syllable durational and distance measures loaded on the third factor. This three-factor solution emerged from each of four conditions involving manipulations of speed of delivery and linguistic context. Canonical correlations showed that the best of four linear additive models involving 3 of the 11 variables was able to explain 45% of the variance in the remaining 8 variables. The three best predictor variables were (a) descending movement displacement, (b) linguo laryngeal movement onset delay and (c) time to peak velocity for the ascending movement. It is argued that these variables represent the speech system's control over the articulatory distinction of different sounds, interarticulatory coordination, and perhaps, rhythm adjustment, and that the first two factors correspond to central variables of breakdown in Broca's aphasia.

One of the classic problems of the empirical investigation of speech has been the selection of maximally informative measurement variables. Such variables should relate to theoretically relevant factors of normal and pathological language performance. As well, they should be reproducible within and across subjects, and should predict as much as possible the behavior of other, related variables.

The selection of maximally informative variables can be explored in the context of a data base of tongue dorsum activity in the vertical plane associated with the production of the syllable /ka/, recently collected by means of ultrasound and simultaneous standard acoustical recordings. The choice of this syllable was in part determined by the limits inherent in the ultrasound instrumentation. For reasons of transducer placement and ultrasound penetration, real-time ultrasound measurement of tongue activity is largely restricted to the imaging of vertical tongue dorsum movement (Keller & Ostry, 1983). Nevertheless, with the appropriate angle, tongue activity that relates to what Ladefoged, Harshmann, Goldstein and Rice (1978) would call "back raising" (page 1027) can be recorded reliably. This is a displacement of the rear tongue mass toward the palate, a distinctive maneuver in the production of the back vowels. Subsequent to recording, correlational analyses of measurements of tongue displacement, duration and movement velocity can be used to assess the relative importance of each of these measurements. Initially, a factor analysis can identify correlations (or a lack thereof) between subsets of measurements. Subsequently, canonical correlations can identify those variables capable of explaining a high percentage of the variance in the remaining variables.

In the present context, five variables were of particular theoretical interest (cf. Figure 1 & Table 1):

1. The amount of descending movement displacement (the vertical displacement from moment 1 to 4) may correspond most closely to the motor-control system's specification of articulatory distinction for different sounds, because smaller values on this variable are typically associated with syllables sounding like/kɔ/ or /kə/ (see Keller, 1986, in press).

2. The descending maximum velocity (the velocity at moment 2) is also of theoretical interest, because it corresponds to stiffness (and thus perhaps muscular stiffness) in a spring- loading model of muscle movement in speech. It has been argued that long and short syllables in stressed and unstressed articulations of /ka/ are distinguished by the slope of the correlation between maximum velocity and displacement (Ostry, Keller & Parush,

1983). This may mean that muscular stiffness contributes to the distinction between these two types of syllables. Further, reduced velocities in articulator movement have been reported for a number of dysarthric conditions (Hirose, Kiritani, & Sawashima, 1982).

3. The linguo laryngeal movement onset delay (LLMOD) (time from moment 1 to 3) is another variable of interest. It is somewhat similar to the traditional measure of voice onset time (VOT), because it is a measure of temporal coordination between the two main articulators involved in the production of /ka/. It has recently been argued that the need for closely timed interarticulatory coordination in speech forms the basis of the ubiquitous temporal constraints acting on speech-motor events (Keller, 1986, in press). This variable is thus expected to have high predictive power for the behavior of a number of other variables.

4. Cycle duration (time from moment 1 to 6) estimates the time between successive occurrences of the syllable /ka/. It therefore corresponds most closely to the time between what Fujimura (1981) would call successive "icebergs" (i.e., major, incompressible articulatory events) in the articulatory stream (see also Fujimura, 1983). It would be of interest to examine if cycle duration predicts the behavior of other durational variables.

5. The time to peak velocity in the ascending movement (time from moment 4 to 5) corresponds to a measurement that we have found to be strongly influenced by the presence of linguistic context (close to invariable in context, highly variable without context). It is possible that the syllable length is adjusted to the requirements of rhythmic articulation of context-free articulation by varying especially the length of the initial portion of the ascending movement. This impression is reinforced by the observation of "platforms" in the smoothed records of long context-free syllables, which are absent in the records of short context-free syllables or syllables produced in linguistic context (marked "A" in Figure 2).

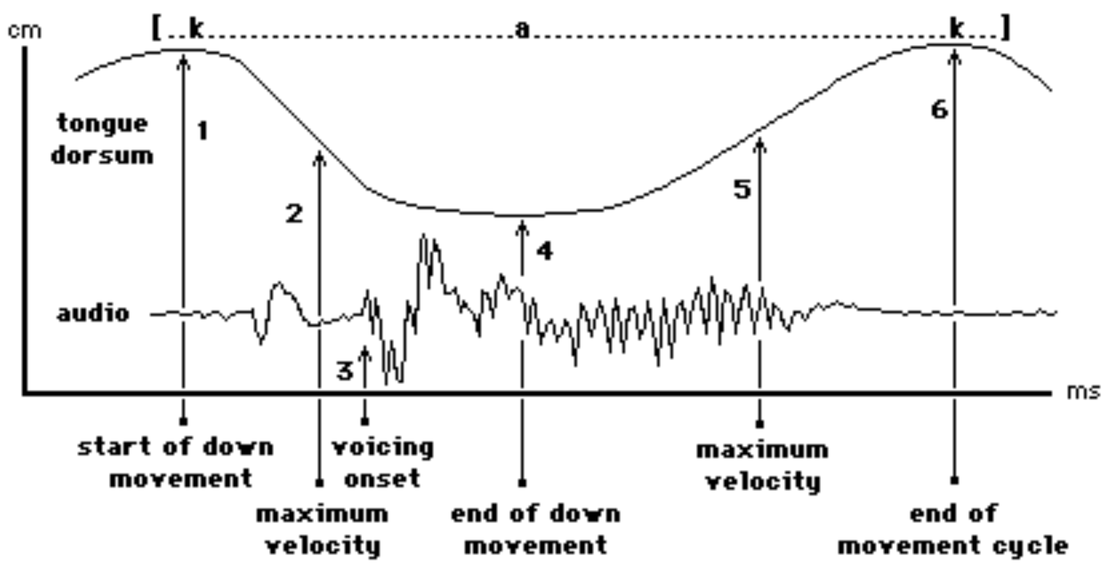

FIGURE 1. The measurement points in the ultrasound and voice trace. The distortion in the voice trace is due to the display system. See Method for the definition of measurement points.

TABLE 1: Measurement variables

	VARIABLE	DEFINITION
1	Displacement, descending movement	distance 1 - distance 4[1]
2	Displacement, ascending movement	distance 4 - distance 6
3	Displacement to peak velocity, descending	distance 1 - distance 2
4	Displacement to peak velocity, ascending	distance 4 - distance 5
5	Cycle duration	time 1 - time 6[2]
6	Duration, descending movement	time 1 - time 4
7	Time to peak velocity, descending	time 1 - time 2
8	Time to peak velocity, ascending	time 4 - time 5
9	Linguo laryngeal movement onset delay (LLMOD)	time 1 - time 3
10	Maximum velocity, descending movement	velocity at time 2
11	Maximum velocity, ascending movement	velocity at time 5

[1]Displacements: a difference of distance x - distance y refers to the vertical distance in cm traversed by the tongue between point x and point y in figure 1. (e.g., distance 1 - distance 4 corresponds to the vertical displacement between points 1 & 4).

[2]Durations: a difference of time x - time y refers to the time in ms elapsed between point x and point y in figure 1. (e.g., time 1 - time 6 corresponds to the duration between points 1 & 6).

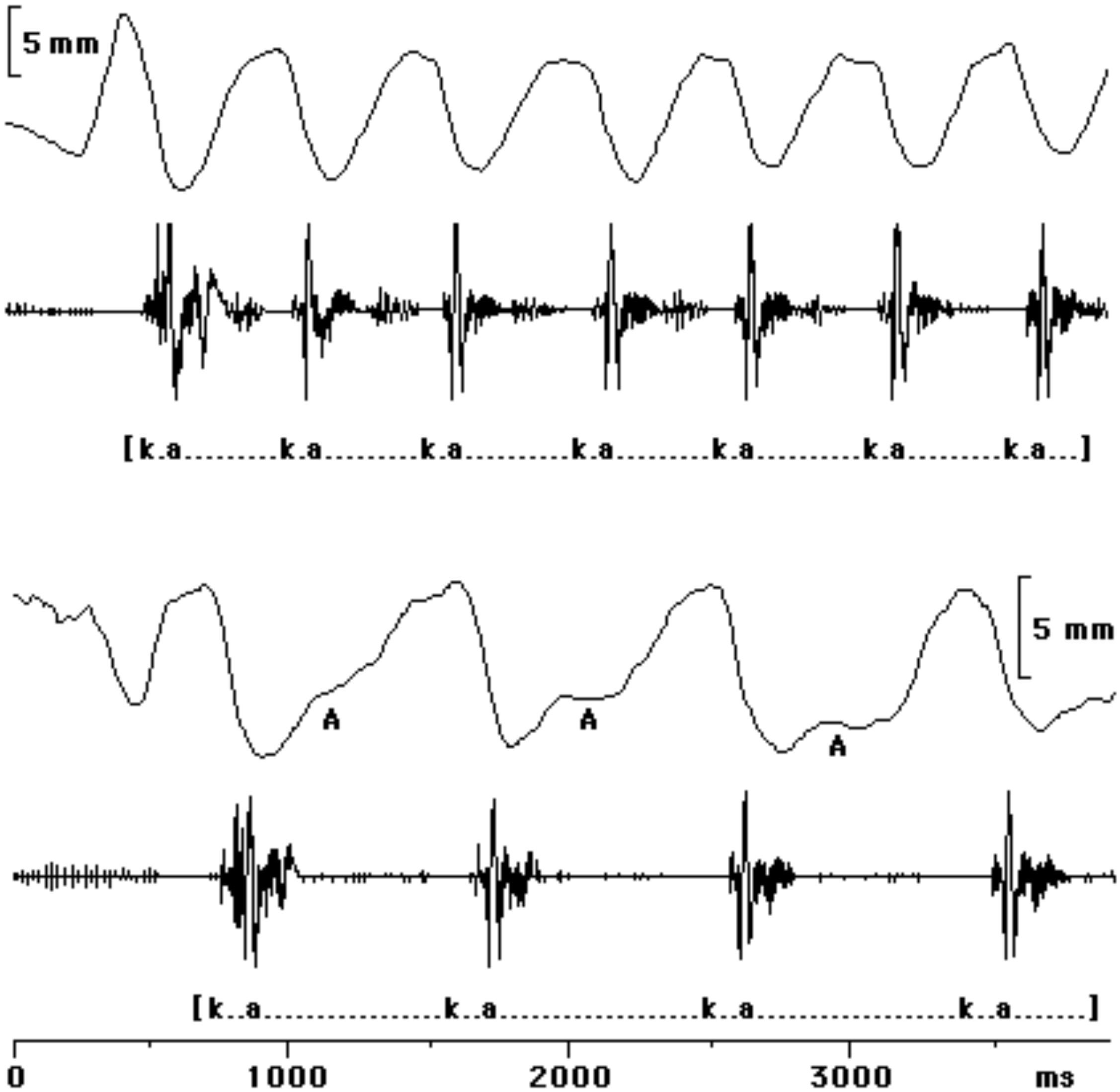

FIGURE 2. Smoothed raw records of a normal francophone speaker repeating /ka/ at rapid pace (above) and at normal, deliberate pace (below). At the beginning of the ascending movement, it is possible to observe "platforms" of syllable lengthening in the normally paced speech (marked A).

A number of these variables have previously been shown to be correlated when measurement for long and short syllables were combined. Ostry et al. (1983), for instance, demonstrated high degrees of correlation between displacement and maximum velocity in isolated repetition of /ka/. However, it remains possible that kinematic relations evident over the full range of long and short syllables may not be identical to those governing a single type of syllable. It was thus one of the purposes of the present study to examine whether such correlations are also found when short and long syllables are examined separately, and when such syllables are embedded in a linguistic context.

METHOD

<u>Subjects and Stimuli</u>

Eleven native speakers of Canadian French, between 20 and 65 years old, 6 female and 5 male, contributed to this study. Each pronounced the syllable /ka/, presented in a balanced random design, at least 25 times in four conditions:

1. /ka/ as a long syllable, obtained in normally paced, context-free repetition [ka:ka:ka:...],

2. /ka/ as a short syllable, obtained in rapid, context-free repetition [kakaka...],

3. /ka/ as a long syllable in the context "le ma<u>ca</u>que assommé" [lə maka:kasɔme],

4. [ka] as a short syllable in the context "le la<u>c à</u> canards" [lə lakakanar].

Prior to recording, subjects repeated the stimuli a number of times to habituate themselves to a smooth delivery, particularly across the morpheme boundary in condition 4. No French context could be found in which the syllable /ka/, embedded between two other /ka/ syllables, does not cross a morpheme boundary. However, once a smooth delivery was established, no phonetic effect of the morpheme boundary could be discerned.

Recording and Data Processing

The recording was performed by a computerized ultrasound measuring device presented in detail in Keller & Ostry (1983). In short, an ultrasound transducer was placed in vertical position below the chin, perpendicular to the Frankfort horizontal line, and was adjusted such that the syllables /ku/, /ko/, and /ka/ could be distinguished in the oscilloscope tracing. The transducer emitted a series of 4 µs 3.5MHz pulses at a 1kHz rate. The echo returning from the tongue dorsum was captured by a peak detector circuit and was interpreted in terms of the momentary distance between the tongue dorsum and the transducer. This information, alongside the 12 bit A/D converted audio signal, was acquired at a 1kHz rate by a laboratory computer. Each recording lasted 4.5 s and a total of 20 recordings were required per subject. If a recording provided less than 95% real data points, it was suppressed online and was repeated. The acquired ultrasound data were smoothed offline by (a) averaging the raw data over 43-ms segments and (b) connecting averages by means of curved lines defined by cubic spline functions. Previous investigations have shown that averaging over 43 ms provides an optimal tradeoff point between considerations of reliability in data imaging and the rejection of high-frequency recording imprecisions (Keller & Ostry, 1983).

Measurement Variables

The measurement points are illustrated in figure 1. Points were marked with a cursor on the display screen, permitting the elimination of deficient records. Point 1 represents the beginning of the descending movement and is defined as the point closest to zero velocity at the peak of the displacement curve. (The velocity trace, not displayed here, is available as the first derivative of the cubic spline functions defining the displacement trace). Point 2 represents the moment of maximum descending velocity, point 3 is the onset of regular oscillations in the audio signal, reflective of the onset of laryngeal activity for the vowel [a], point 4 is the end of the descending movement and the beginning of the ascending movement (velocity closest to 0), point 5 is the moment of maximum ascending velocity, and point 6 represents the end of the ascending movement, and at the same time, the end of the syllable /ka/ (velocity closest to 0). In all measurements making reference to points on the velocity curve (all points except point 3), the rightmost measurement was chosen if several measurements were similarly close to zero or to maximum velocity. For point 3, the first glottal pulse in the acoustic voice trace was defined as the first regular pulse deviating from the background noise observed during the preceding linguo palatal closure.

Eighteen measurements of displacement, duration, and maximum velocity were derived from the six measurement points, of which 11 non redundant measurements were selected for the correlational analyses (Table 1). As an example of a redundant variable ineligible for multiple correlational analysis, ascending movement duration (time from moment 4 to moment 6) is redundant, once descending movement duration (time from moment 1 to moment 4) and syllable duration (time from moment 1 to moment 6) have been selected for analysis.

Data Analysis

To render data from the 11 subjects comparable, measurements were converted to standard (z) scores and outliers were suppressed. About 3% of observations were eliminated through outlier suppression (values exceeding 3 SD from the mean). In each condition, subjects contributed equal numbers of observations to the data base. The total Ns in the combined database were 385 (11 x 35) for contextual long /ka/, 374 for contextual short /ka/, 198 for context-free long /ka/, and 396 for context-free short /ka/. For each measurement variable in each condition, the probability that data were not drawn from a normal population was < .05. None of the linear correlations between individual variables exceeded .799, and most were in the range of .2 to .6. Factor analyses and canonical correlations were run separately for each condition.

Table 2: Factor loadings

	In linguistic context, long (N=385)			In linguistic context, short (N=374)		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Displacement, descending movement	.880*	- .178*	- .037	.825*	- .322*	- .026
Max. velocity, descending movement	.846*	.135*	- .020	.824*	- .081	.006
Displacement, ascending movement	.817*	- .113	- .376*	.814*	- .057	- .310*
Max. velocity, ascending movement	.808*	- .076	- .169*	.794*	.119	- .112
Displacement to peak velocity, descending	.680*	- .413*	.269*	.659*	- .414*	.247*
Displacement to peak velocity, ascending	.520*	- .097	- .770*	.455*	.002	- .823*
Duration, descending movement	.110	- .895*	- .084	.094	- .849*	.020
Cycle duration	.127	- .835*	- .259*	.174*	- .759*	- .249
Time to peak velocity, descending	.192*	- .831*	.208*	.206*	- .674*	.266
Linguo laryngeal mov. onset delay (LLMOD)	- .074	- .826*	- .058	- .080	- .806*	- .015
Time to peak velocity, ascending	- .022	- .051	- .906*	- .086	- .003	- .902*
Eigenvalues	3.616	3.123	1.778	3.389	2.700	1.794
Percent of explained variance	32.871	28.393	16.162	30.813	24.543	16.307
Cumulative percent of explained variance	32.871	61.264	77.426	30.813	55.356	71.663

	Context-free, long (N=198)			Context-free, short (N=396)		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Displacement, descending movement	.843*	- .205*	- .060	.847*	- .216*	- .250*
Max. velocity, descending movement	.794*	.106	.025	.809*	.093	- .169*
Displacement, ascending movement	.807*	- .064	- .184*	.791*	- .161*	- .414*
Max. velocity, ascending movement	.665*	.035	- .074	.782*	- .129*	.074
Displacement to peak velocity, descending	.544*	- .426*	.036	.677*	- .351*	- .087
Displacement to peak velocity, ascending	.537*	.102	- .668*	.493*	- .094	- .749*
Duration, descending movement	.156	- .825*	- .047	.215*	- .858*	- .116
Cycle duration	- .078	- .510*	- .613*	.208*	- .677*	- .496*
Time to peak velocity, descending	.076	- .869*	.010	.153*	- .838*	.012
Linguo laryngeal mov. onset delay (LLMOD)	- .100	- .882*	- .037	- .007	- .850*	.021
Time to peak velocity, ascending	.053	.027	- .941*	.044	- .030	- .947*
Eigenvalues	3.068	2.725	1.757	3.426	2.851	1.993
Percent of explained variance	27.887	24.773	15.969	31.145	25.916	18.117
Cumulative percent of explained variance	27.887	52.660	68.629	31.145	57.061	75.178

*Significance level (two-tailed test): $p<.01$.
Loadings exceeding .5 are italicized.

RESULTS

Factor Analysis

When the number of factors after varimax rotation was limited by the criterion of Eigenvalue ≥ 1.0, the analysis provided three factors for each condition (table 2). Variables loading in excess of .5 on factor 1 were movement displacements and maximum velocities. Variables loading at >.5 on factor 2 were all measurements of duration. The two variables loading at >.5 on factor 3 were time to peak velocity in ascending movements and the associated displacement. This three-factor solution explained between 68.6 and 77.4% of the total variance (M = 74.7%). Factor loadings were similar in the four conditions.

A plot of loadings on factor 1 versus loadings on factor 2 reinforces the impression that the data incorporate three dimensions (figures 3 & 4). Displacements and maximum velocities are clearly separate from durations, and time to peak velocity for the ascending movement is separate from either of the previous sets of points. The membership of the displacement to peak velocity in the ascending movement is arguable, because factor loadings place it midway between the displacements and the time to peak velocity for the ascending movement.

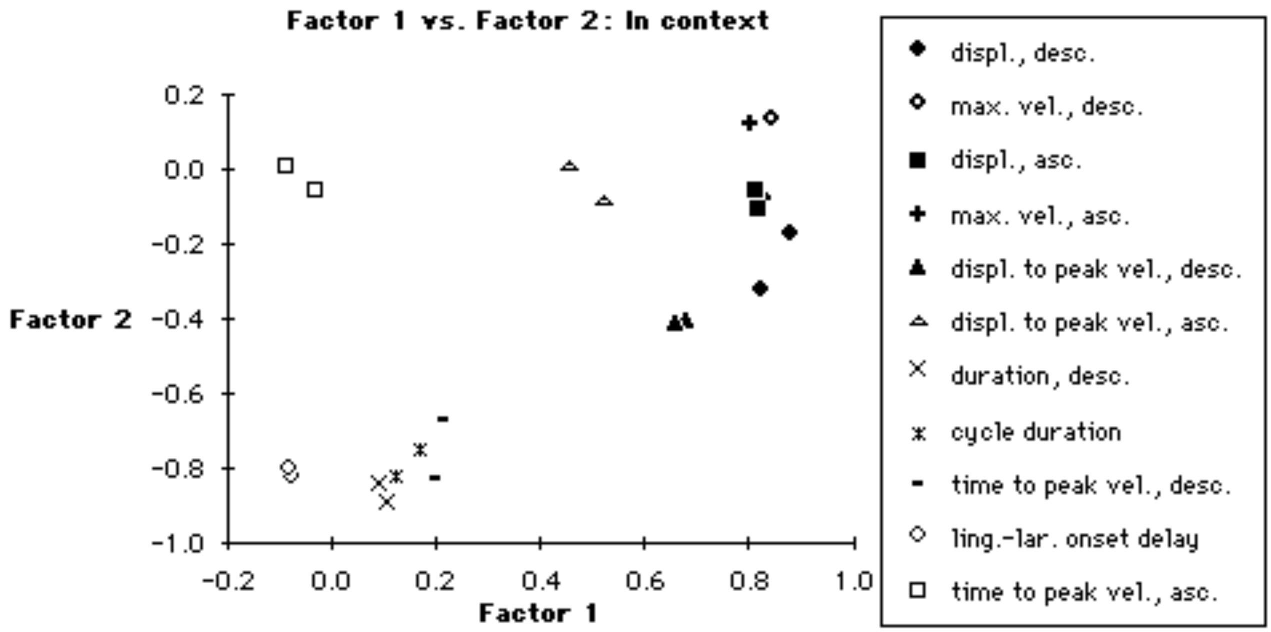

FIGURE 3. A plot of Factor 1 versus Factor 2 in the conditions with linguistic context. Variables loading heavily on Factor 1 (distances and velocities) are found predominantly in the upper right corner, variables loading heavily on Factor 2 (durations) are predominantly in the lower left corner, and variables loading heavily on Factor 3 (time and distance to peak velocity in the ascending movement) are in the upper left and top of the graph.

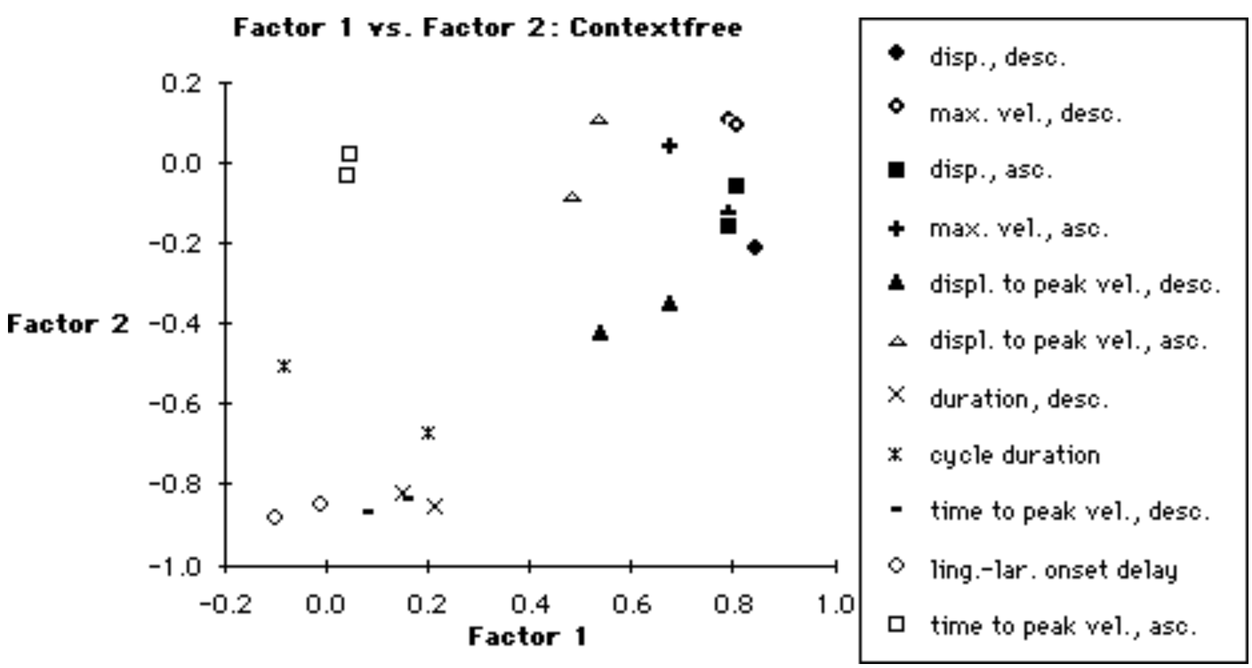

FIGURE 4. A plot of Factor 1 versus Factor 2 in the context-free conditions. Plot distributions are very similar to those in Figure 3.

<u>Canonical Correlations</u>

To permit further meaningful variable selection, four competing linear-additive models were examined by means of canonical correlations. The four models involved the five predictor variables of theoretical interest mentioned in the Introduction. All five variables loaded primarily on just one of the three factors just described.

1. Descending movement displacement + linguo laryngeal movement onset delay + time to peak velocity, ascending	predict values on the remaining 8 variables.
2. Descending movement displacement + cycle duration + time to peak velocity, ascending	predict values on the remaining 8 variables.
3. Descending maximum velocity + linguo laryngeal movement onset delay + time to peak velocity, ascending	predict values on the remaining 8 variables.
4. Descending maximum velocity + cycle duration + time to peak velocity, ascending	predict values on the remaining 8 variables

The analysis of the four competing predictive models revealed a marked superiority of the first model, because it was able to account for the greatest average adjusted variance in the remaining eight variables (Table 3). Specifically, canonical correlations of the type "$\text{Variable}_{1...8}$ = Descending movement displacement + linguo laryngeal movement onset delay + time to peak velocity, ascending" produced the highest adjusted R^2 values, averaged over $\text{variable}_{1...8}$. These three variables alone can predict about 45% of the variance in the remaining variables, while the other three combinations were able to account for just 35 to 42% average adjusted variance. This model explained, in descending order, the following adjusted variances, averaged over the four conditions: Maximum descending movement velocity, .617; displacement to peak velocity, ascending, .593; descending movement duration, .517; ascending movement displacement, .483; cycle duration, .399; time to peak velocity, descending, .361; displacement to peak velocity, descending, .349; maximum ascending movement velocity, .292.

This model was also tested with values from a single subject. A 32 year old man provided 120 observations in all four conditions, after separate z scoring and outlier suppression were performed for each condition. Model 1 explained on the average 48.5% of the variance in the eight remaining variables, model 2 explained 46.0%, model 3 explained 40.2% and model 4 explained 46.8%. Although the advantage of model 1 is less clearcut in this single subject (perhaps because data from the four conditions were combined or because of the small n), the first model did command the lead and explained close to half the variance in the data.

DISCUSSION

A factor analysis of 11 measures of lingual activity involved in the production of the syllable /ka/ has indicated that three factors can explain about 75% of the variance in the data. Variables loading strongly on the first factor were displacement and velocity measures, variables loading strongly on the second factor were durational measures, and variables loading strongly on the third factor related to durational and distance measures of the mid portion of the syllable. The stability of this three-factor solution was demonstrated by the fact that the four task conditions did not induce any systematic differences in factor loadings. Also, the correlational patterns for separate long and short syllable conditions with and without linguistic context are in general accord with correlational patterns reported for combined long and short, context-free syllables reported in Ostry et al. (1983).

A further analysis contrasting four possible predictor models by means of canonical correlations has shown that a linear-additive model with 3 of the 11 variables was able to explain on the average 45% of the variance in the remaining 8 variables. These three predictor variables were (a) descending movement displacement, (b) linguo laryngeal movement onset delay, and (c) time to peak velocity for the ascending movement.

This result may be cautiously interpreted in the following speech motor-control terms. One of the speech control system's primary responsibilities is to perform articulatory gestures that produce perceptually categorized, distinct sounds. To distinguish [ka] from other possible syllables such as [kə] or[kɔ], this probably means specifying distinctively greater descending movement displacement values than would be required for either of the other syllables (in addition to information in other planes not captured in the present data). The good predictive power of descending movement displacement is in accord with this notion, because it indicates that the behavior of a number of variables correlates with the degree of vertical displacement achieved in downward displacement.

Another major responsibility of the system is to assure that articulations are performed in conformity with constraints on interarticulatory timing. In the syllable /ka/, one important interarticulatory timing constraint is the delay between the onset of lingual and laryngeal movements. The prominence of linguo laryngeal movement onset delay may be a reflection of such temporal constraints.

The interpretation of the variable most predictive of the third factor, time to peak velocity in the ascending movement, is relatively uncertain. One clue might be differences in variability of this measurement, which pilot analyses have related to the presence or absence of linguistic context. In these analyses, time to peak velocity in the ascending movements has been found to be close to invariable in linguistic context, but highly variable in conditions without linguistic context. It is possible that the motor-control system adjusts the syllable length to the requirements of the rhythmic articulation of a diadochokinetic production of [ka:ka:ka:...] by varying particularly the length of the initial portion of the ascending movement. By contrast, articulation within the given linguistic contexts may have been relatively less rhythmical and may have required less adjustment on this variable. This possibility will be further explored in another study.

It is interesting to note that some of the most prominent speech motor-control perturbations in Broca's aphasia can be characterized in terms of the first two of these factors. Broca's aphasics are noted for their phonemic substitution errors (e.g., Blumstein, 1973; Keller, 1978), which may correspond to misspecifications of movement distance (factor 1). This type of error is much less frequently found in nonmotoric types of aphasia. Also, Broca's aphasics are known for their impairments in interarticulatory coordination (cf. problems with VOT documented by Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980 and by Itoh, Sasanuma, Tatsumi, Murakami, Fukusako, & Suzuki, 1982); this type of problem may relate to misspecifications of factor 2. By contrast, nonmotoric aphasics do distinguish VOTs of voiced and unvoiced sounds, even though their distributions of this temporal measurement deviate from the norm.

A final comment concerns maximum velocity. The present analysis did not provide much support for a central role of descending maximum velocity, contrary to possibilities explored previously (Ostry et al., 1983). The factor analysis indicated that this variable covaries primarily with displacement variables, and the canonical correlations showed that it does not have a particularly high predictive power in explaining the variance of other measurement variables. Further, a careful inspection of the relevant figures in Ostry et al. (1983, p. 630) reveals that displacement was in fact a better predictor of membership in the stressed and unstressed syllable groups than maximum velocity. Despite its great theoretical interest, the present analysis indicates no particularly prominent role for maximum velocity in explaining the variance of related variables.

Table 3: Average percent of adjusted variance explained in remaining eight variables by four sets of predictors

PREDICTORS	In linguistic context long /ka/	In linguistic context short /ka/	Context-free long /ka/	Context-free short /ka/	M
Desc. mov. displ. + LLMOD + time to peak vel., asc.	.480	.417	.405	.503	.451
Desc. mov. displ.+ cycle duration + time to peak vel., asc.	.497	.399	.311	.464	.418
Desc. mov. max. vel. + LLMOD + time to peak vel., asc.	.416	.372	.352	.255	.349
Desc. mov. max. vel. + cycle duration + time to peak vel., asc.	.468	.375	.255	.423	.380

Note: Highest values are italicized. All solutions are significant at $p<.01$ (Wilk's lambdas ranging from .012 to .154).

ACKNOWLEDGEMENTS

This research was supported by the National Sciences and Engineering Council of Canada, by the Fonds FCAR (équipes) of the Province of Québec and by the Programme d'aide financière aux chercheurs et aux créateurs of the Université du Québec à Montréal. I wish to thank my research assistants Ginette Ladouceur and Yvan Barbeau for their excellent work in data recording and analysis.

REFERENCES

Blumstein, S. (1973). *A phonological investigation of aphasic speech.* The Hague: Mouton.

Blumstein, S.E., Cooper, W.E., Goodglass, H, Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language*, *9*, 153-170.

Fujimura, O. (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica*, *38*, 66-83.

Fujimura, O. (1983). *A linear model of speech timing.* Murray Hill, NJ: Bell Laboratories.

Itoh, M., Sasanuma, S., Tatsumi, I.F., Murakami, S, Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain and Language*, *17*, 193-210.

Hirose, H., Kiritani, S., & Sawashima, M. (1982). Velocity of articulatory movements in normal and dysarthric subjects. *Folia Phoniatrica*, *34*, 210-215.

Keller, E. (1978). Parameters for vowel substitutions in Broca's aphasia. *Brain and Language*, *5*, 265-285.

Keller, E. & Ostry, D. (1983). Computerized pulsed echo ultrasound measurements of tongue dorsum movements. *Journal of the Acoustical Society of America*, *73*, 1309-1315.

Keller, E. (1986). The cortical representation of motor processes of speech. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language.* (pp. 125-162). Hillsdale, NJ: Lawrence Erlbaum Associates.

Keller, E. (in press). Analyse fonctionnelle des perturbations neurologiques de la parole. In H. Cohen (Ed.), *Perspectives contemporaines en neurosciences.* Montréal: Études Vivantes.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America*, *64*, 1027-1035.

Ostry, D., Keller, E. & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology, Human Perception and Performance*, *9*, 622-636.