Dynamic aspects of English vowels in /bVb/ sequences

Keith Johnson

Abstract

Analysis of X-ray microbeam recordings of 5 speakers pronouncing /bVb/ sequences revealed that vowels in midwestern American English differ from each other in terms of lip, tongue, and jaw dynamics as well as in terms of "target" positions. The data supported Lehiste & Peterson's (1961) distinction between short and long-nucleus vowels. Short-nucleus vowels /ɪ, ɛ, ʌ/ had shorter deceleration phases during the lip opening movement and shorter acceleration phases during the lip closing movement. A kinematic analysis of the consonant opening and closing movements suggested that in a spring/mass model of articulator movement these short vowels would be characterized by greater spring stiffness. /ɔ/ had less spring stiffness during the opening gesture and /o^{U}, e^{I}, æ/ and /ɔ/ had less spring stiffness during the closing gesture. A canonical discriminant analysis of articulator positions across time found consistent patterns of tongue movement which separated the vowels into the same groups found in the kinematic analysis of lip movement. Each of the four factors found in the analysis was associated with a movement pattern. Additionally, the first two factors were associated with gross tongue location differences and the third factor was associated with tongue bunching. These analyses suggest that the dynamic control of the lip gestures in /bVb/ sequences is coordinated with tongue movement patterns for vowels.

1. Introduction

In addition to spectral distinctions (Peterson & Barney, 1952), American English vowels differ in acoustic duration and formant trajectory patterns. Low vowels such as the vowel in "hod" have longer durations than do high vowels such as the vowel in "heed", and the vowels in "hid", "head", "hood", and "hud" have shorter durations, respectively, than those in "heed", "heyed", "who'd", and "hod" (Lehiste & Peterson, 1961). Lehiste & Peterson also noted that the vowels in "hid", "head", "hood", and "hud" had relatively shorter F2 steady-states than did the other vowels. Therefore, they classified /ɪ, ɛ, ʊ/ and /ʌ/ as short-nucleus vowels, /i, u, e^{I}, o^{U}, ɑ, ɔ, æ, a^{U}, a^{I}/ and /o^{I}/ as long-nucleus vowels. The long-nucleus vowels could be further divided into simple- and complex-nucleus vowels, and among the complex-nucleus vowels, Lehiste & Peterson identified /e^{I}, o^{U}/ and /ɚ/ as single-target vowels and /a^{U}, a^{I}/ and /o^{I}/ as double-target vowels. The double-target vowels typically had two F2 steady-states, while the others did not. Stevens, House & Paul (1966) found that vowels produced in CVC sequences in which the initial and final consonants were identical did not have symmetric F2 trajectories. So, for example, the F2 trajectory of [gu] was not a mirror image of the F2 trajectory of [ug]. They reasoned that these formant trajectory asymmetries reflected the existence of offglides in the vowel. They suggested, for example, that /i/ and /u/ had peripheral offglides and could be transcribed [i^{j}] and [u^{w}], while Lehiste & Peterson's (1962) short-nucleus vowels had central offglides and could be transcribed [V^{ə}]. The distinction between simple and complex nucleus vowels has also been a feature of linguistic descriptions of American English since at least 1933 (Bloomfield, 1933, p. 104, 124; Trager & Smith, 1951, p. 12 ff.; Chomsky & Halle, 1968) although the particulars of each successive analysis vary.

There is increasing evidence that listeners expect and make use of dynamic information in vowel perception (Strange, 1989). Huang (1986) and DiBenedetto (1989a,b) found that the temporal location of the peak in the F2 trajectory has an impact on the categorization of some American English vowels. Strange, Jenkins & Johnson (1983) found that listeners' vowel identification performance was only slightly affected when the vowel centers in CVC syllables were replaced by silence. Parker and Diehl (1984) confirmed this finding. When the middle 70% of the vowel was replaced by silence, vowel identification error rates were about 20% for short vowels and less than 10% for long vowels (the comparable error rates in the full vowel condition were 5% and 3%). Even with 90% of the vowel replaced by silence (only about 10-15 ms of the vowel onset and 10-15 ms of the vowel offset remaining audible) Parker & Diehl found that listeners' performance was well above chance. These results were taken to indicate that vowel formant transitions provide valuable perceptual information, which listeners readily use. This conclusion was strengthened by Verbrugge & Rakerd's (1986) silent-center vowel study. They constructed silent center vowel stimuli by splicing together initial (or final) transitions taken from vowels produced by men, with final (or initial) transitions produced by women (and an appropriate amount of silence between the two portions). They found that these hybrid male/female silent center vowels were identified just as accurately as were single-speaker silent center vowels. This result suggests that 'target' formant frequencies may be less important in vowel perception than are the direction and rate of change of formant trajectories, and thus that spectral change in an important perceptual property for vowels in English. Nearey and Assman (1986) also came to this conclusion. They constructed stimuli from naturally produced isolated vowels by extracting a 30 ms portion from early in the vowel and another 30 ms portion from late in the vowel. Listeners could correctly identify the vowels when these vowel portions were played in the original order, but if the first portion was played twice or the two portions were played in the opposite order the listeners' performance dropped dramatically.

The two-target representation proposed in linguistic descriptions of American English fits several aspects of these acoustic and perceptual studies of vowels. First, some of the duration differences among American English vowels may reflect a difference between vowels with two vowel targets and vowels with one vowel target. Similarly, Lehiste & Peterson's (1961) distinction between short-nucleus and long-nucleus vowels can be described in terms of the number of articulatory targets involved in producing the vowel. Note, however, that this is not Lehiste & Peterson's (1961) interpretation. They distinguished between three types of long-nucleus vowels only one of which, in their view, had two vowel targets (the "true" diphthongs [a^{I}, a^{U}] and [o^{I}]). We will return to this point in the conclusion. A two-target model of vowel articulation in American English is obviously relevant for Nearey and Assman's (1986) perceptual study, and may also provide an explanation for the relative importance of vowel edges as opposed to vowel centers found in the silent center studies reviewed by Strange (1989). In addition, a phonetic distinction between one and two-target vowels corresponds to a distinction which must be made in view of some phonological phenomena. For instance, phonetically long vowels may occur in open syllables such as "bee", "bay", "spa", "law", "go", "do", and [bæ] (the noise a sheep makes), and in open upbeat syllables with secondary stress such as in the words "recede", "Daytona", "tautology", "rotation", "bubonic" and "Camay" while phonetically short vowels may not occur in these environments.

In contrast to the acoustic and perceptual studies which suggest that changes in formant trajectories during American English vowels are linguistically significant, many previous studies of vowel production have focussed on articulatory target positions during vowels and various sources of variability for these targets assuming that a vowel in American English is specified by a single articulatory target. For instance, Kent & Netsell (1971) reported the effects of linguistic stress on tongue, jaw and lip positions at the acoustic midpoints of vowels with some illustrative data on articulatory dynamics of stress distinctions. Kent & Moll (1972b) studied vowel-to-vowel coarticulation, reporting movement trajectories from one vowel target to another with various consonants or linguistic boundaries intervening. Ladefoged, DeClerk, Lindau & Papçun (1972) studied individual differences in tongue shapes for vowel targets, and their data were further analyzed in terms of tongue shape factors by Harshman, Ladefoged & Goldstein (1977). Gay (1974) noted the effects of consonant / vowel coarticulation, vowel-to-vowel coarticulation, and speaking rate on the positions of the tongue, lips and jaw at the point of maximum articulator displacement during vowels. Perkell & Nelson (1982) investigated variability in tongue positioning during vowel production as a function of the place of maximal constriction, looking at one time slice for each vowel (the point of extreme movement toward the vowel target). Jackson's (1988) cross-linguistic study was based on tongue shape data taken at the vowel midpoint. All of these studies have in common that they characterize the articulation of American English vowels in terms of a single vowel target.

So, acoustic phonetic studies indicate that the vowels of English have discernable dynamic properties and perceptual studies indicate that these dynamic properties are important for speech perception. Also, the traditional linguistic analysis of American English vowels (going back to Bloomfield, 1933) suggests that the vowels can be separated into those which have two vowel targets and those which have only one. Yet, the dynamics of vowel articulation in American English have not been extensively studied. The experiment reported here addresses this issue by analyzing (1) the kinematics of lower lip movement for different vowels in /bVb/ sequences as produced by speakers of American English, and (2) tongue body movements during those vowels.

2. Method

The data were collected by Peter Ladefoged and Mona Lindau at the x-ray microbeam facility at the University of Wisconsin (Fujimura, Kiritani & Ishida, 1973; Kiritani, Itoh & Fujimura, 1975; Abbs, Nadler & Fujimura, 1988). Some aspects of these data have been reported previously (Lindau & Ladefoged, 1989,1990; Johnson, Ladefoged & Lindau, submitted).

2.1 *Subjects*

Five speakers (3 females and 2 males) of northern midwestern American English served as speakers for the experiment. They were paid a small sum for their participation and were recruited by the staff at Wisconsin from the university community. The subjects were unaware of the specific purposes of the experiment, and reported no history of speech or hearing deficiencies and had no dental fillings. The speakers were screened for dialect homogeneity by having them read several sets of dialect diagnostic words (e.g. "merry", "Mary", and "marry"). One of the male speakers (RP) did not distinguish between /ɔ/ and /ɑ/, so he did not read the /ɔ/ words. For these speakers, /æ/ was diphthongized and could be transcribed as [ɛ^ə] or [e^ə], and /ɛ/ was transcribed as somewhat lower than in other dialects of American English.

2.2 *Materials*

The speakers read sentences containing symmetric C_iVC_i sequences with the consonants /d, b, s/ and the vowels /i, ɪ, e^ɪ, ɛ, æ, ɑ, ɔ, ʌ, o^ʊ, ʊ, u/. Not all of these sequences were real words in English and so the subjects were instructed in the pronunciation of the non-English sequences by pointing out words which rhyme with the test sequences. For instance, "beb" [bɛb] rhymes with "Deb" in this dialect. In an attempt to balance the demands of using actual words with the desire for a factorial experimental design, some of the words had CVC structure and some had CV structure with the following word in the carrier phrase supplying the final C of the sequence. For example, instead of being asked to read, "Say beeb (/bib/) between", the subjects read, "Say be between". The different syllable structures complicated the interpretation of lip closing kinematics as discussed below. A full list of the sentences is presented in Table I. Analyses of the /bVb/ sequences are reported below. This subset of utterances was chosen on the assumption that in them tongue movement would be relatively free from consonant effects and thus more easily interpretable as reflecting vowel articulations (Engstrand, 1988; Nord, 1975).

2.3 *Procedure*

Small (2.5 mm) gold pellets were glued to the speakers' lips, teeth, and tongue along the midline of the vocal tract (Figure 1). Two additional pellets tracked head movement. These pellets were glued to the bridge of the nose and to the border of the upper incisor and gums. One pellet was glued to the border of the lower incisors and the gums and indicated the location of the jaw. The lip pellets were glued to the borders of the vermilion ridges of the upper and lower lips. The tongue pellets were placed at intervals of approximately 15 mm on the protruded tongue with the

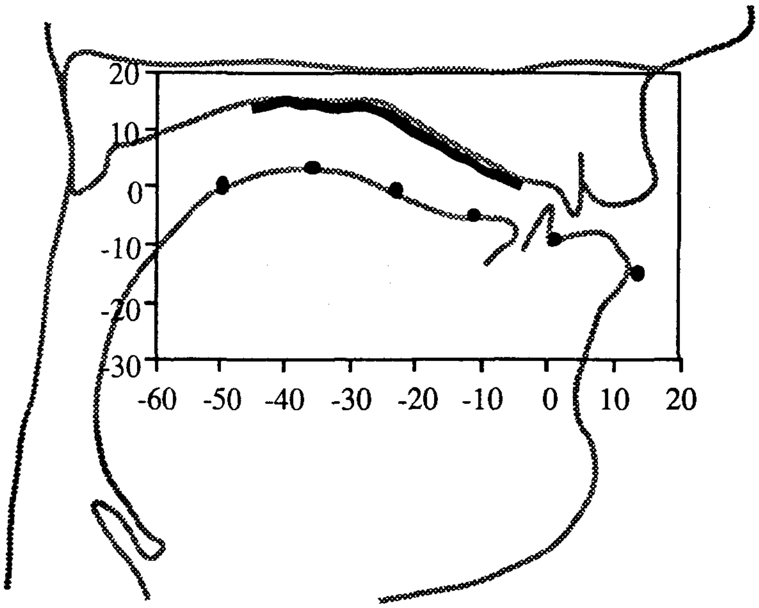
Table I
List of materials.

Say	dee	to me.	Say	bee	between.	Say	see	serenely.
	did			bib			sis	
	day			bay			say	
	dead			beb			cess	
	dad			bab			sass	
	Dodd			bob			soss	
	daw			baw			saw	
	doe			boe			sew	
	dood			----			soos	
	do			boo			sue	
	dud			bub			suss	

first pellet about 8-10 mm behind the tongue tip. Figure 1 shows the locations of the pellets at the midpoint of the vowel averaged across all vowels, consonants, and speakers and shows that when the tongue was not protruded the pellets were on average about 10 mm from each other. As the talkers read the experimental materials the movements of the pellets were tracked by a computer controlled x-ray system (Nadler, Abbs & Fujimura, 1987). A small beam of x-ray tracked each pellet, and the locations of the pellets (in both the vertical and horizontal dimensions) were recorded at intervals of 10 ms (for tongue, lower lip and nose) or 20 ms (for jaw and upper lip). Accuracy of the measurements was on the order of fractions of a millimeter. Additionally, the speech wave form was simultaneously sampled at a rate of 10 kHz.

Each sentence was repeated three times in a given recording run and the entire procedure was performed twice by each subject giving a total of six repetitions of each CVC sequence per subject. Thus, the total number of possible utterances was 960. Of this number 300 contained /bVb/ sequences and only 202 utterances (67% of the total) were available for analysis due to various types of experimental error. The statistical analyses therefore were based on unequal numbers of observations of the different vowels. Most of the missing observations were due to missing data collection runs, therefore for some of the tokens for some subjects only three observations per vowel were available.

After the data had been collected, the nose and upper incisor pellet traces were used to correct for head movements, rendering the other pellet traces in terms of movement relative to the speaker's occlusal plane rather than absolute movement. Five events were located in the two dimensional movement trajectory of the lower lip for each CVC sequence (Figure 2): (1) the point of maximum displacement toward consonant closure during the initial consonant, (2) the point of maximum speed (the change in displacement in two dimensions per unit time) from consonant closure to vowel opening , (3) the point of maximum vowel opening, (4) the point of maximum speed from vowel opening to the final consonant closure and (5) the point of maximum displacement toward consonant closure during the final consonant. A computer program located these articulatory events for each vowel utterance and recorded (1) the times of the events, (2) the locations of the pellets at each event, (3) the magnitudes of the opening and closing gestures, and (4) the peak speed values of the opening and closing gestures. This program has been previously described (Johnson, Ladefoged & Lindau, submitted) and will be briefly summarized here. For each utterance, the time of the acoustic onset of the vowel had been previously identified by eye in a digital wave form display and stored in a computer file. The measurement program looked at the lower lip trajectory in a window of time around the acoustic onset of the vowel and found the locations of maximum displacement and speed. The trajectories were evaluated in two dimensions, so maximum consonant displacement was defined as the point at which the lower lip was furthest



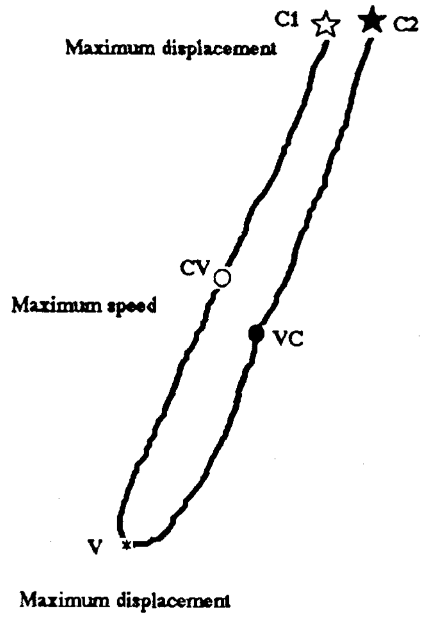


Figure 1(on the left). Placement of the pellets on the surfaces of the vocal tract.
Figure 2 (on the right). Schematic representation of the articulatory landmarks at which pellet locations were measured. This figure represents the locations of maximum displacement and speed of the lower lip. C1 = point of maximum displacement during the initial /b/. V = point of maximum displacement during the vowel. C2 = point of maximum displacement during the final /b/. CV = point of maximum speed during the opening movement. VC = point of maximum speed during the closing movement.

up and forward and the maximum vowel displacement was defined as the point at which the lower lip was furthest down and back. The time of the maximum vowel displacement measured in this way was not reliably different from previous measurements made from visual displays of the lower lip vertical movement (Lindau & Ladefoged, 1990).

3. Lower lip movement

This section describes vowel effects in the movement trajectories of the lower lip in /bVb/ sequences. Each trajectory can be summarized in terms of the durations of the component movements and the displacement amplitudes and peak velocities of the lower lip.

Figure 3 shows the relative times of the articulatory landmarks illustrated in Figure 2 for each vowel averaged across speakers. Each opening movement is composed of an acceleration phase (C1 to CV) and a deceleration phase (CV to V). Similarly, each closing movement is composed of an acceleration phase (V to VC) and a deceleration phase (VC to C2). The vowel nucleus can be defined as the portion of the trajectory extending from CV to VC. Based on the data in Figure 3 the vowels of English can be divided into two classes; short-nucleus /ɪ, ɛ, ʌ/ and long-nucleus vowels /i, e^{I}, æ, ɑ, ɔ, o^{U}, u/ (see Lehiste & Peterson, 1961 for acoustic evidence for this distinction, and that /ʊ/ is also a short-nucleus vowel). The short-nucleus vowels had shorter overall durations and also shorter nuclei. Within short-nucleus vowels, the opening acceleration (C1 to CV) was longer than the opening deceleration (CV to V) and the closing deceleration (VC to C2) was longer than the closing acceleration (V to VC). The opposite pattern occurred in long-nucleus vowels (longer opening deceleration and longer closing acceleration). Across vowels, the duration of the opening acceleration was correlated with movement amplitude and did not separate the vowels into short versus long-nucleus. The duration of the opening deceleration, on the other hand, was not correlated with movement amplitude; the short-nucleus vowels having shorter opening decelerations than the long-nucleus vowels. The closing decelerations of short-nucleus vowels were longer than those of the long-nucleus vowels. This seemed to be true regardless of syllable structure; "bab" and "Bob" both had short closing decelerations as did "baw", "bow", "boo", etc. rather than long closing decelerations as in "bib", "beb", and "bub". The long opening deceleration for /ɔ/ seems to have been the result of was more lip rounding early in the vowel than late [ɔ^{ə}], pushing back the point of maximum lip opening (V). Conversely, the maximum displacement (V) during /o^{ʊ}/ was also not in the center of the vowel nucleus, suggesting greater rounding at the end of the vowel than at the beginning.

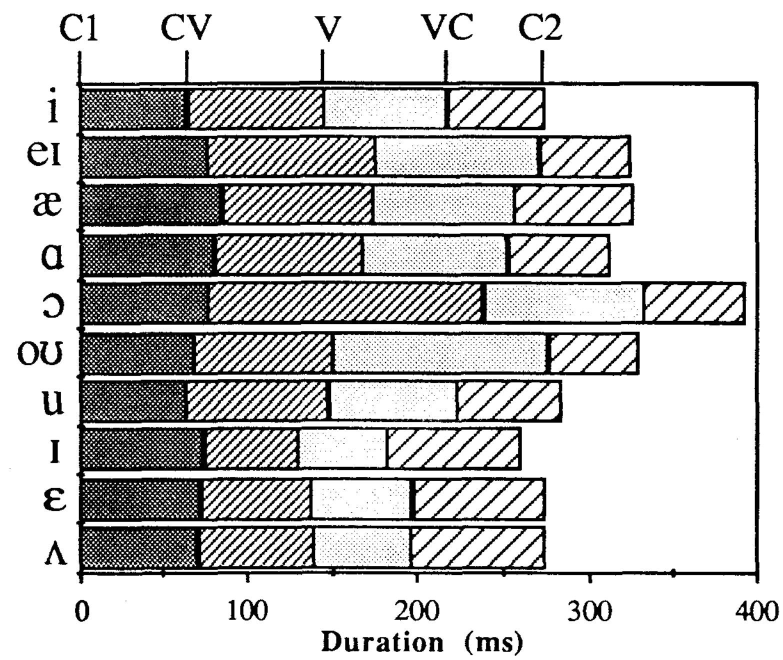


Figure 3. Average durations of lower lip trajectories during /bVb/ sequences. The labels at the top of the figure refer to the articulatory landmarks illustrated in Figure 2. The opening gesture extends from C1 to V, the closing gesture from V to C2. The opening acceleration is C1 to CV, the opening deceleration is CV to V. The closing acceleration is V to VC, and the closing deceleration is CV to C2.

Figure 4 shows kinematic data from the opening and closing movements. The vowel symbols in this figure represent the average values for that vowel. Following Beckman, Edwards & Fletcher (1991) these data can be interpreted in terms of a spring/mass dynamic system (Saltzman & Munhall, 1989). The positive correlation between opening displacement amplitude and peak speed (left panel) agrees with Kent & Moll's (1972a) observation that the further an articulator must move the faster the movement, and suggests that the opening movements differ primarily in terms of an underlying amplitude parameter for the opening gesture. In addition, the distinction between /ɔ/ and the other vowels appears to involve articulator stiffness. /ɔ/ lies below the regression line indicating that on average its opening movement was accomplished more slowly than was the opening movement for other vowels with similar displacement amplitudes. In a spring/mass model this pattern can be simulated by reducing spring stiffness. The short-nucleus vowels /ɪ, ɛ, ʌ/, on the other hand, can be characterized as having *greater* spring stiffness than the others. /ɪ, ɛ, ʌ/ lie above the regression line in the left panel of Figure 4. So, the duration differences illustrated in Figure 3 can be interpreted in terms of a spring/mass dynamic model. Differences in the duration of the opening acceleration were associated with differences in movement amplitude in the model (which is at odds with a model in which the vowels are distinguished solely by gestural amplitude), and differences in the duration of opening deceleration were associated with spring stiffness in the model.

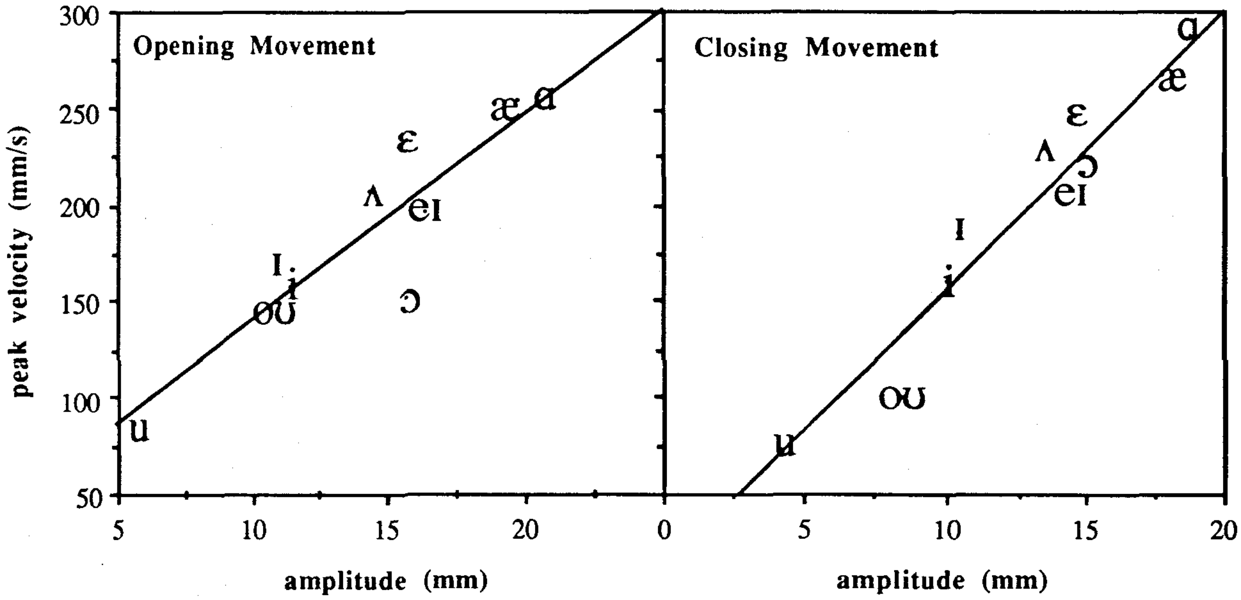


Figure 4. Opening movement and closing movement kinematics. Each symbol represents the values for the indicated vowel averaged across speakers. The regression lines were calculated from the vowel averages.

The right panel of Figure 4 shows kinematic data from the closing movements. As before, the vowels appear to differ primarily in terms of gestural amplitude, but also as before there appear to be some interesting stiffness differences. Like the opening movement of /ɔ/, the closing gesture of /o^{ʊ}/ appears to be characterized by lower stiffness; it had low speed relative to its displacement amplitude. To a lesser degree this seems to be the case also for /e^{ɪ}, æ/ and /ɔ/. The closing gestures of these three vowels all had lower velocities than other vowels having about the same displacement amplitudes (i.e. they lie below the regression line in the right panel of Figure 4). In addition, the closing movements of the short-nucleus vowels had higher velocities than other vowels having about the same displacement, and so could be characterized as having greater stiffness in a spring/mass model. Whereas it was possible to associate changes in stiffness with the duration of the opening deceleration and changes in gestural amplitude with the duration of the opening acceleration, the relationship between the kinematic model parameters of the closing gestures and the durations of the closing movement's components is not obvious. Increased stiffness for /ɪ, ɛ, ʌ/ was associated with longer closing decelerations and shorter closing accelerations, while decreased stiffness for /o^{ʊ}, e^{ɪ}, ɔ, æ/ tended to be associated with longer closing accelerations. Gestural amplitude was not correlated with total closing gesture duration nor with either of the movement phases. The interpretation of these data is complicated by the fact that syllable structure is a confounding variable in this study (some sequences having CV structure and some having CVC structure). Also, there may have been some aspects of tongue movement which constrained the lip gestures and placed vowel-dependent constraints on their temporal structure. We turn now to this topic.

4. Vocal tract configurations

This section describes the results of a factor analysis of articulatory movements in /bVb/ sequences. The acoustic and perceptual studies mentioned in the introduction suggest that dynamic information may be important in maintaining vowel distinctions. The factor analysis was designed to identify reliable patterns of movement during the vowels. The locations of six pellets at three times during each vowel were entered into a canonical discriminant analysis (Kshirsagar, 1972; SAS Institute, 1982). The analysis gave a derived vowel space and factor loadings for each of the dimensions of the derived space. These factor loadings can be translated back into the articulatory space and can be interpreted as abstract articulatory patterns involved in vowel production.

The average locations of the pellets at three times are shown in Figure 5. The legend in this figure and others to follow refers back to Figure 2. The three times are the points of maximum lower lip speed during lip opening (CV) and lip closing (VC), and the point of maximum lower lip displacement during the vowel (V). The four pellets on the tongue will be referred to (from front to back) as tongue tip, tongue body 1, tongue body 2, and tongue dorsum. Figure 5 shows that, averaged across vowels and speakers, the lower lip showed a displacement of about 9 mm from CV to V, and about 6 mm from V to VC. The jaw showed similar directions of movement with much smaller magnitudes and the tongue pellets reflected the same pattern with decreasing magnitudes further back in the mouth. The average tongue dorsum pellet location showed almost no movement averaged over the vowels. The carrier phrase was "say /bVb/ between", with the vowel [eᴵ] preceding and [ɨ] following the test word. Figure 5 suggests that vowel-to-vowel coarticulation did not produce a unique movement pattern of the tongue in this context. The tongue movements indicated in the figure appear to be due solely to the movement of the jaw which presumably was coordinated with the lower lip in producing the bilabial closures. It is not clear whether the asymmetry in the lip positions was due to the context in which the words occurred or whether this pattern would be observed in any context.

Figure 6 shows the average pellet positions for the ten vowels. As in other studies of vowel articulation (Stevens & House, 1955), the vowels were separated from each other by differences in the location and degree of vocal tract constriction at the center of the vowel. The two back tongue pellets were higher during /æ/ than during /ɛ/, while the lip opening was greater for /æ/. As mentioned earlier, /æ/ for these speakers was impressionistically transcribed as [ɛᵊ].

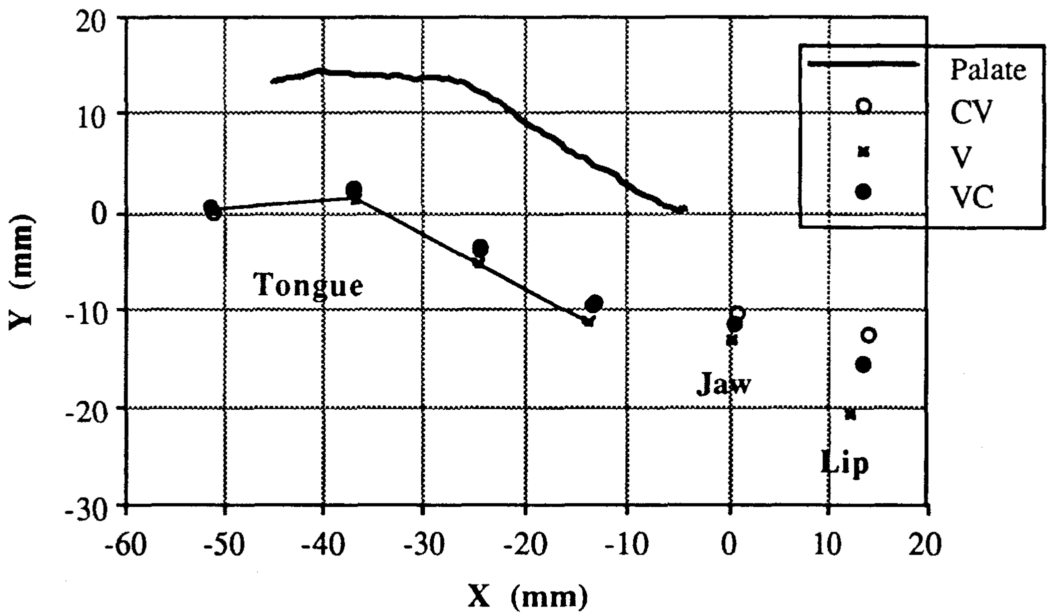


Figure 5. Pellet locations at three times averaged across speakers and vowels. The legend refers to the articulatory events illustrated in Figure 2.

Figure 6 also indicates that there were significant movement patterns associated with several of the vowels. /e^{I}/ and /ɔ/ had tongue raising and some tongue fronting during the vowel. /æ/ had tongue retraction and lowering, /o^{U}/ and /u/ had tongue retraction and raising, and /ɛ/ and /ɪ/ had some tongue lowering.

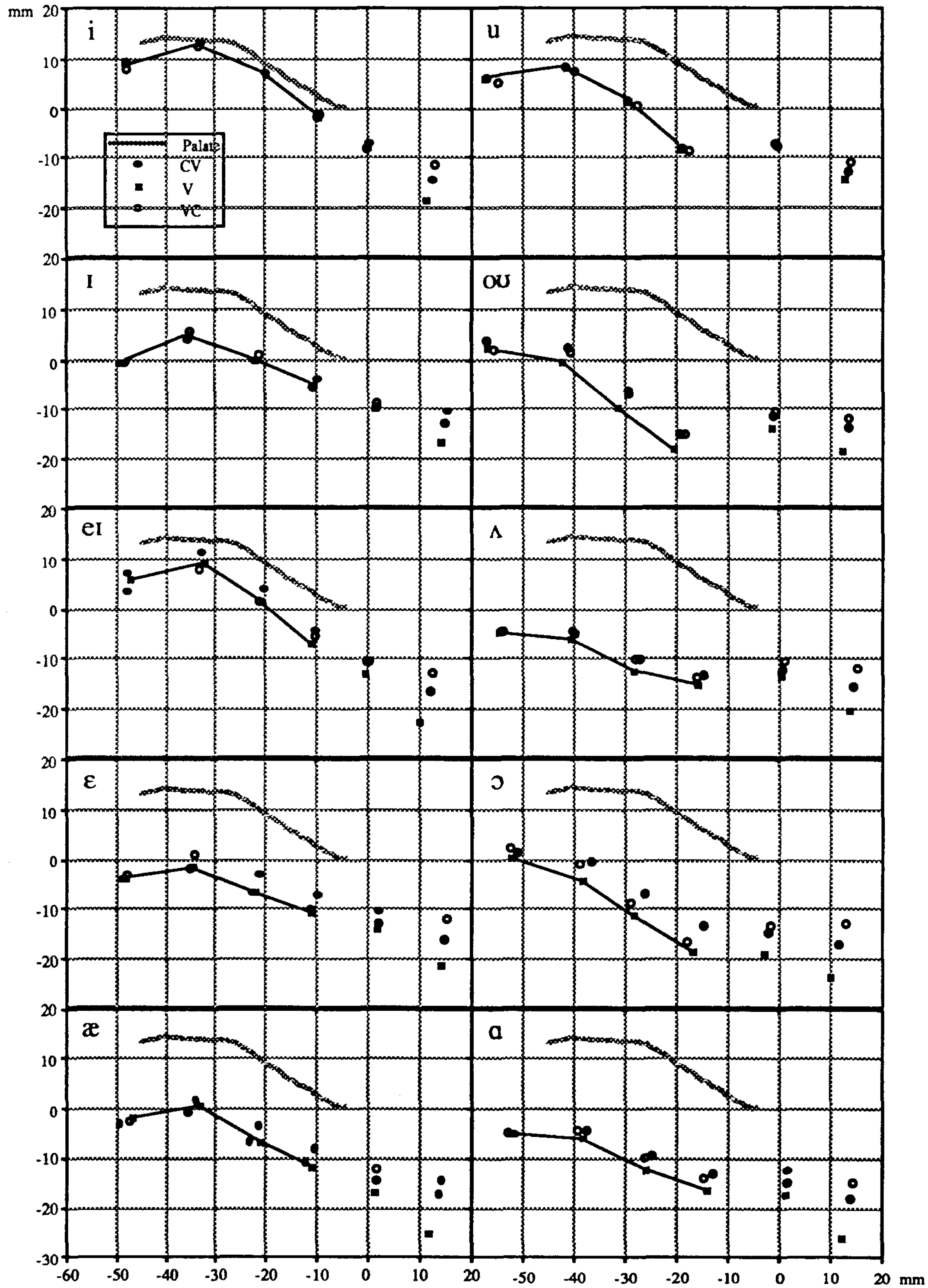


Figure 6. Pellet locations at three times averaged across speakers. Each panel shows the average pellet locations for a particular vowel.

Rather than rely on subjective impressions about the data in Figure 6, canonical discriminant analysis was used to explore general patterns of vocal tract posture and movement. Canonical discriminant analysis finds principal components along which categories can be best discriminated. "Given a classification variable and several quantitative variables, canonical discriminant analysis derives canonical variables (linear combinations of the quantitative variables) that summarize between-class variation in much the same way that principal components summarize total variation" (SAS Institute, 1982). In the analysis reported here, the classification variable was vowel identity and the quantitative variables were the pellet locations at three times. Thus for each of the 202 observations there were (6 pellets X 2 dimensions X 3 times =) 36 quantitative variables and one classification variable. The analysis found coefficients for a set of equations of the form:

(1) $$v_j = x_1 a_{1j} + x_2 a_{2j} + \ldots + x_{36} a_{36j}$$

where v_j is the value of canonical variable j, the x_n are the 36 quantitative variables, and the a_{nj} are the canonical coefficients for canonical variable j. For each canonical variable the coefficients are optimized to produce maximal separation between categories (in this case vowels). After finding the first set of coefficients (the set which accounts for the greatest amount of between category variability), coefficients for another canonical variable, uncorrelated with the first, are found. This procedure is repeated N-1 times where N is the number of categories. Because the input quantitative variables are transformed to make the pooled within-class covariance matrix an identity matrix, the canonical variables do not represent perpendicular components through the space of the original variables (a common complaint about principal component analysis). The average scores for a particular vowel on the canonical variables (the v_js) define that vowel's location in the derived vowel space.

The relationship between the canonical variables and the original quantitative variables can be calculated in the following way. First, a scale factor for canonical variable j (c_j) is calculated by (2). Where r_{ij} is the correlation between canonical variable j and quantitative variable i, and b_{ij} (derived from a_{ij}) is the standardized canonical coefficient on canonical variable j for quantitative variable i. The patterns of variation (x_{ij}) in the original quantitative variables which are encoded by the canonical variables are then given by (3). Where sd_i is the standard deviation of quantitative variable i and ave_i is the average value of quantitative variable i. If in formula (3) we set v_j equal to a large positive value (within the range of observed values for v_j), the x_{ij} give the pattern of values on the original quantitative variables associated with positive values of canonical variable j. This procedure was used to calculate the articulator loadings shown in Figures 9 through 12. We derive predicted values of the quantitative variables for a particular vowel by setting the v_j in (3) equal to the observed v_j for that vowel and summing the z-score component of (3) over j before multiplying by the standard deviation and adding the mean (4). For the sake of continuity with previous research, the canonical variables will be called "factors".

(2) $$c_j = \sum_{i=1}^{36} r_{ij} b_{ij}$$

(3) $$x_{ij} = (r_{ij} v_j / c_j) sd_i + ave_i$$

(4) $$\text{predicted}_i = \sum_{j=1}^{n} (r_{ij} v_j / c_j) sd_i + ave_i$$

Factor analysis has been used previously to identify basic tongue shapes in vowels. Harshman, Ladefoged & Goldstein (1977) found that tongue shapes of the vowels of English could be described with just two factors. Jackson (1988) found that similar tongue shape factors underlie vowel production in English and Icelandic. The main difference between these earlier analyses and the one reported here is that lip and jaw position data and data from 3 times during each vowel were included in the analysis. Consequently, the articulator loadings in this analysis represent underlying patterns of articulator movement which distinguished the vowels in these particular utterances.

The first two factors together accounted for about 80% of the variance, and with the addition of two more factors 90% of the variance was accounted for. The normal method of determining the number of factors to include in a model is to look for an elbow in the variance accounted for curve. By this metric only the first two factors in the present analysis would be chosen. However, since a canonical discriminant analysis gives a unique solution (i.e. the first factor is the same regardless of how many other factors one cares to look at) and the amount of variance accounted for by the third and fourth factors was fairly large (6.9% and 6% respectively, both $p < 0.001$), I will discuss the vowel space and articulatory loadings for the first four factors.

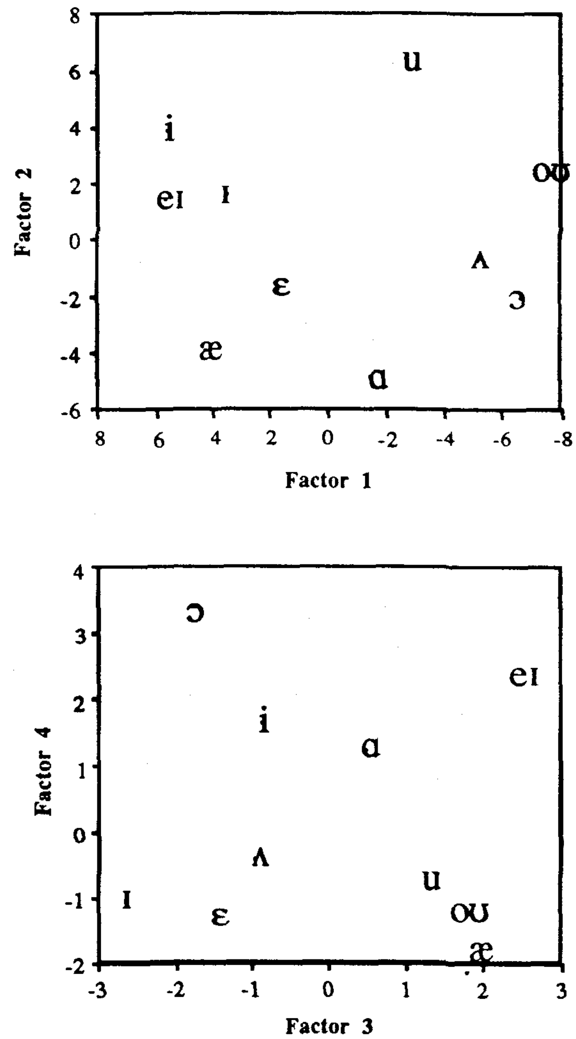


Figure 7. Derived vowel space from the cannonical discriminant analysis. Top: Factor 1 versus factor 2. Bottom: Factor 3 versus factor 4.

Figure 7 shows the four-factor vowel space. The vowel space formed by the first two factors (top panel) is strikingly similar to the traditional impressionistic vowel space. The first factor (the horizontal axis of the top panel) separated the front vowels /i, ɪ, eᴵ, ɛ/ and /æ/ from the back vowels /u, oᵁ, ʌ, ɔ/ and /ɑ/, and the second factor corresponded to the traditional high/low distinction, with /u/ and /ɑ/ having the most extreme values. Note that the short-nucleus vowels were less peripheral in this space than were the long-nucleus vowels, indicating that they were produced with less extreme versions of the first two factors.

The articulatory factor loadings for the first factor (Figure 8) confirmed that vowels with positive values on the factor (top panel) had fronter tongue positions than vowels with negative values on the factor (bottom panel). For instance, the tongue dorsum pellet was about 48 mm behind the upper incisor pellet (the origin of the coordinate space) when factor 1 had a large positive value (top panel), and was about 56 mm back when factor 1 was negative. The tongue shape encoded by the first factor is similar to Harshman et al.'s (1977) "front raising" factor for tongue shapes. Both in Harshman et al.'s analysis and in the present analysis, the first factor distinguished between vowels which had a point of maximum constriction in the front of the mouth (alveolar or post-alveolar) with vowels which had a point of maximum constriction in the pharynx (this must be infered for the x-ray microbeam data, see Lindau & Ladefoged, 1989 concerning this inference). Jackson (1988) also found a front raising component in the production of Icelandic vowels. There are some differences in the detailed tongue shapes found by Harshman et al. (1977), Jackson (1988) and the present analysis, but the general characteristics of the solutions are in agreement.

There are two types of data represented in the present analysis which were not included in Harshman et al. (1977) or Jackson (1988). These are (1) lip and jaw positions and (2) movement over time. Comparison of the jaw pellet in the upper and lower panels of Figure 8 reveals that there was very little difference in jaw position associated with factor 1 (especially as compared with the differences in jaw position associated with the articulatory pattern for factor 2, Figure 9). So, the tongue positions associated with factor 1 differed relative to an essentially fixed jaw. This observation implies that the vowels /eᴵ/ and /oᵁ/ (which differed on factor 1 but not on factor 2) had essentially the same degree of jaw opening while having differing tongue positions. Figure 6 verifies this prediction of the analysis. The lower lip position associated with positive values of factor 1 is relatively low (as compared to the position of the jaw) while the lower lip position associated with negative values of factor 1 was higher and more protruded from the jaw. This indicates negative values of factor 1 were associated with rounded lips.

Figure 8 also shows movement patterns associated with the first factor. The movement patterns found in the factor analysis include vowel and consonant components. Therefore, in evaluating the movement patterns derived in the factor analysis it is necessary to keep in mind the average pattern of movement (Figure 5) for vowels in the /bVb/ context. Some of the tongue movements, particularly of the tongue tip pellet, appear to be jaw related, however, because the tongue dorsum pellet showed no movement in the average pattern, any movement of this pellet in the factor loadings indicate vowel related movements unambiguously. When we compare the movement patterns associated with the first factor to the average pattern of movement, it is apparent that front tongue positions (top panel) were associated with a small degree of tongue backing and tongue body rotation, whereas back tongue positions (bottom panel) were associated with some tongue body and front tongue raising during the closing movement. Generally, however, the degree of within-vowel tongue movement associated with factor 1 was quite small.

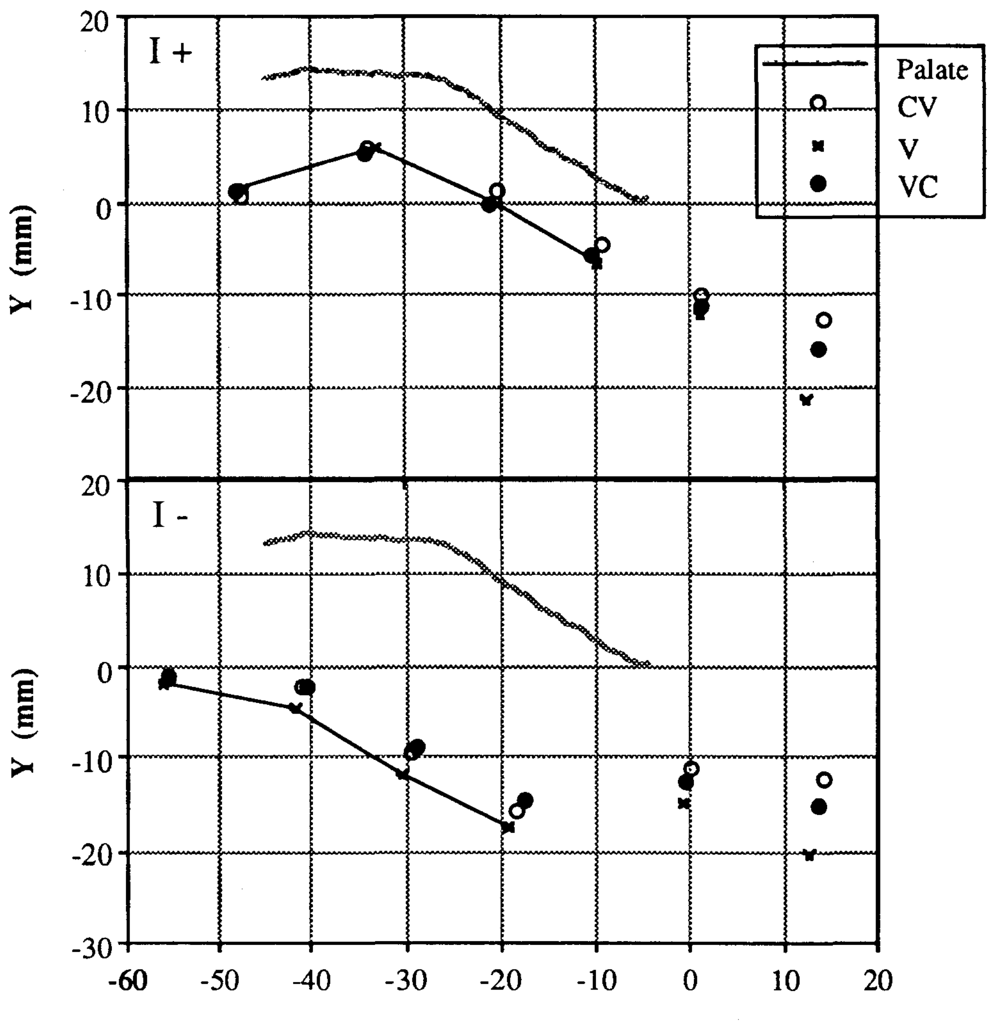


Figure 8. Factor loadings for the first factor. Top: positive loading. Bottom: negative loading.

The articulatory factor loadings for the second factor (Figure 9) had high tongue positions for positive factor values (top panel) and low tongue positions for negative factor values (bottom panel). The tongue loadings on this factor were similar to Harshman et al.'s (1977) "back raising" factor for tongue shapes. Jackson (1988) also found a similar factor for tongue shapes in Icelandic vowels. As was mentioned above, factor 2 was associated with a large difference in the position of the jaw, positive values of the factor were associated with close jaw positions (throughout the vowel) and negative values were associated with open jaw positions. The similarity between factor 2 and Harshman et al.'s (1977) back raising suggests that very open jaw positions (found in the present study) are associated with constriction low in the pharynx (found by Harshman et al.).

As with factor 1, there appears to be a difference in lip rounding associated with factor 2. Positive values of the factor (top panel, Figure 9) were associated with relatively higher and more protruded lip positions at the center of the vowel. Note however that negative values of factor 2 were associated with relatively high lip positions at the vowel edges (CV and VC, in the bottom panel of Figure 9). This may reflect a strategy for attaining a low jaw position in /bVb/ context. If the jaw movement begins before the lip movement (i.e. the lips are help closed while the jaw begins its decent) the amount of time available for jaw movement is greater than it would be otherwise. An adjustment in the relative timing of jaw and lip movement (phase angle change) seems to be indicated by the pattern of movement found for factor 2. The jaw showed no movement for positive values of factor 2, while the tongue moved up and back (particularly during the opening phase of the lip movement). The articulatory loadings for negative values of the second factor were associated with downward and slightly forward tongue movements during the opening phase.

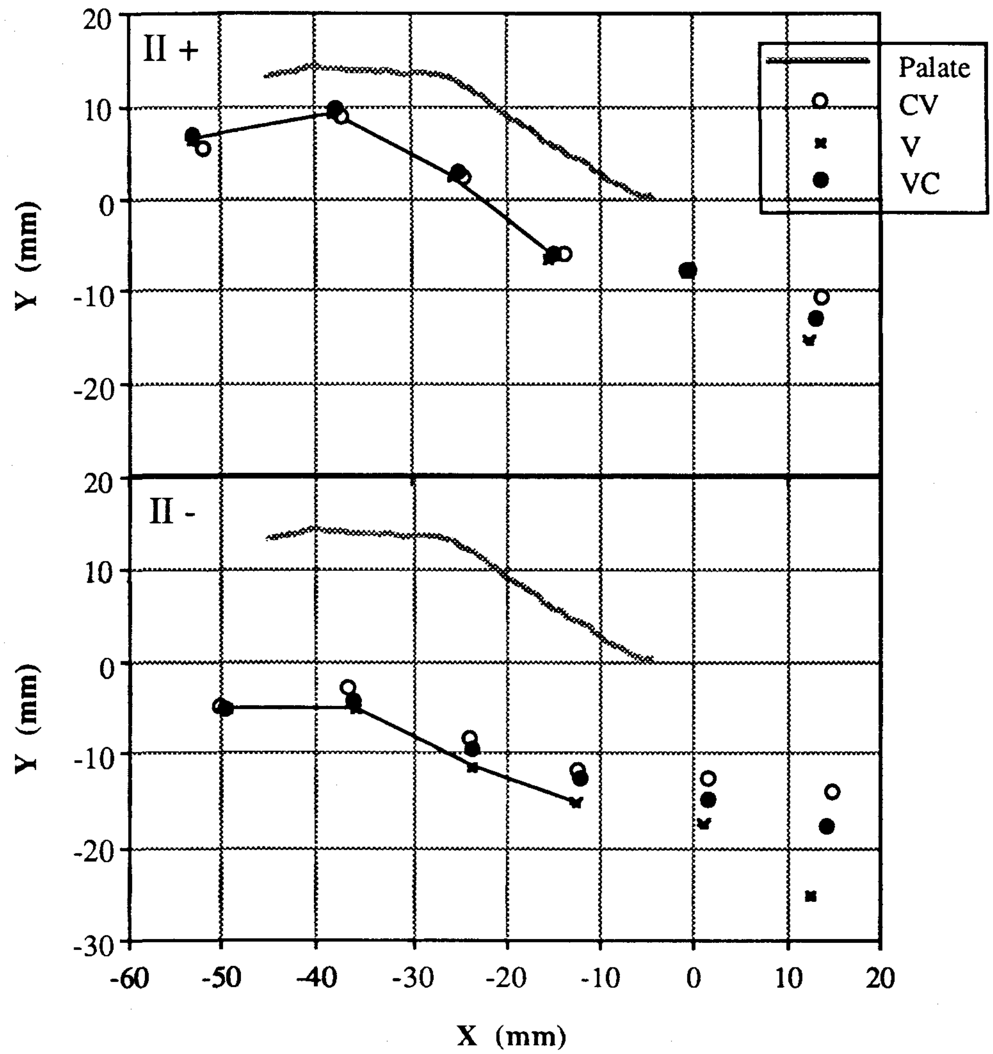


Figure 9. Factor loadings for the second factor. Top: positive loading. Bottom: negative loading.

The bottom panel of Figure 7 shows the average values of the vowels on the third and fourth factors. There are some interesting correspondences between this figure and the lip kinematic data discussed in the previous section. In the lip kinematic data, /ɪ, ɛ/ and /ʌ/ were characterized by stiffer opening and closing lip movements than the other vowels and in Figure 7 they are also separated from the other vowels by having negative values on factors 3 and 4. Both the lip opening and closing movements for /ɔ/ in /bVb/ sequences had reduced stiffness and /ɔ/ was separated from the other vowels by having a large positive value for factor 4 and a negative value for factor 3. /o$^{\text{U}}$, e$^{\text{I}}$/ and /æ/ had less stiff closing movements than the other vowels and also had the largest positive values on factor 3. So, apparently there is some relationship between the kinematic properties of lower lip movement during the vowels and the articulatory patterns found in the factor analysis.

Figure 10 shows the articulatory factor loadings for the third factor. Whereas the first two factors involved large changes in the position of the tongue, the third and fourth factors were associated with more subtle aspects of articulation. The third factor was associated with tongue shape; positive values (top panel) had a more bunched shape than negative values (bottom panel). Also, positive values of the third factor were associated with upward and backward movement of the tongue and negative values were associated with slightly more downward and forward movement than in the average movement pattern (Figure 5). Figure 11 shows the articulatory factor loadings for the fourth factor. As with the third factor, a subtle tongue shape difference was associated with the fourth factor; the tongue dorsum pellet was lower when the fourth factor took a negative value. Positive values on the fourth factor (top panel) were associated with forward and upward movement of the tongue. Negative values of the fourth factor (bottom panel) were associated with a small amount of backward tongue movement and downward movement during the lip opening movement.

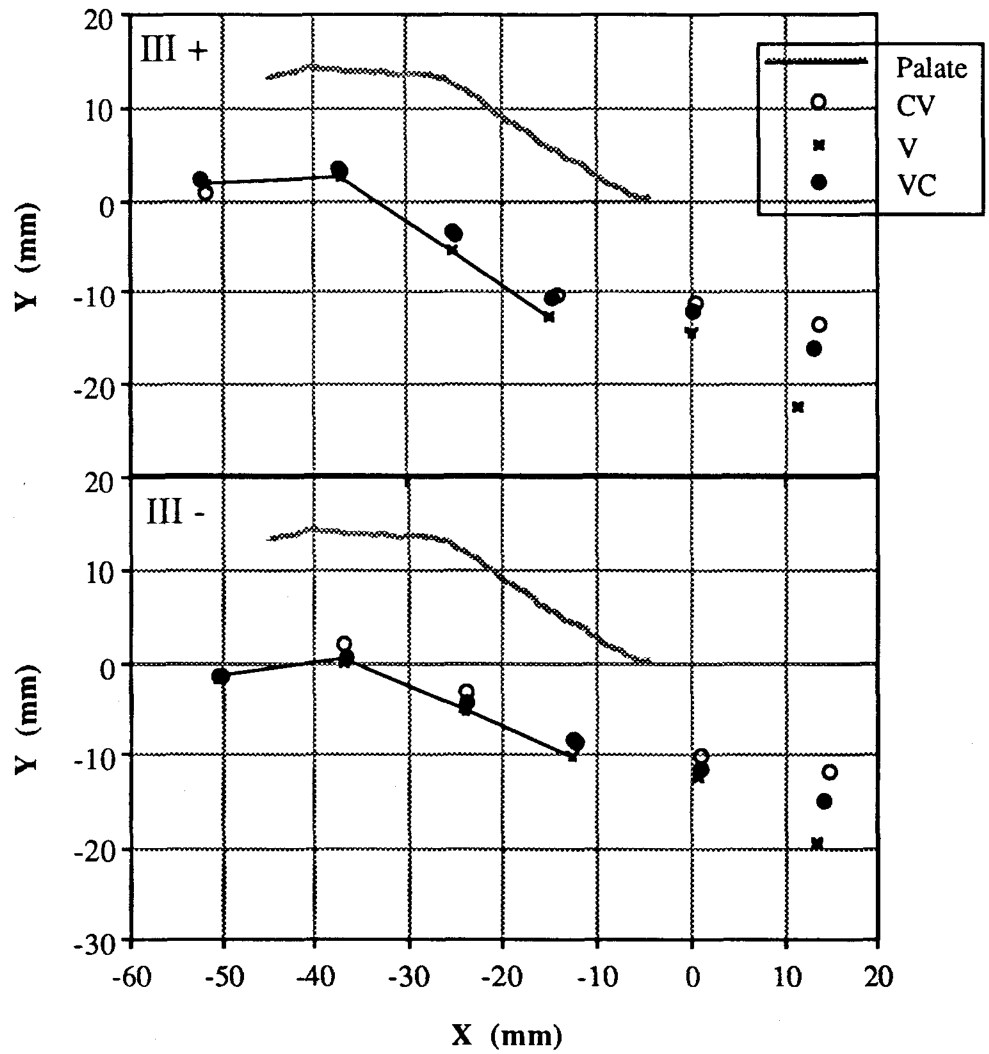


Figure 10. Factor loadings for the third factor. Top: positive loading. Bottom: negative loading.

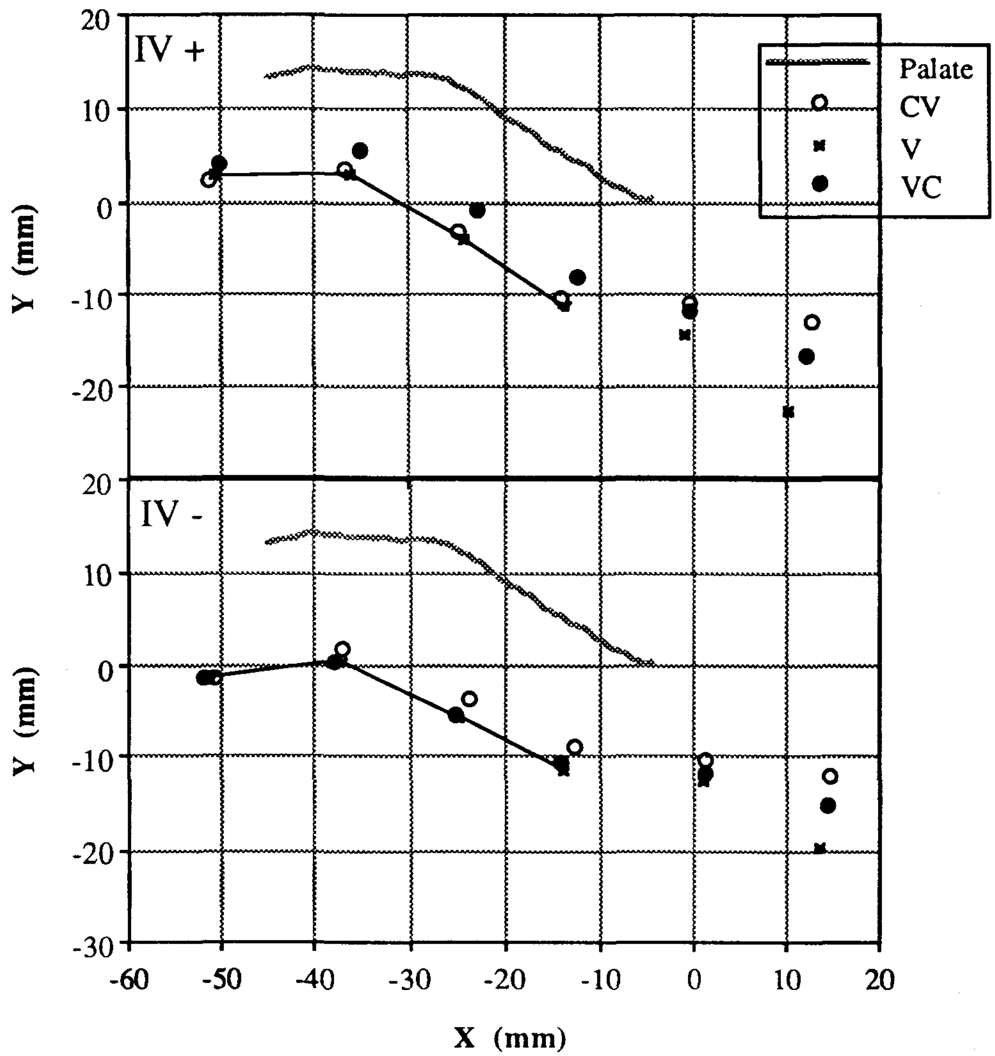


Figure 11. Factor loadings for the fourth factor. Top: positive loading. Bottom: negative loading.

If we focus on the tongue dorsum pellet in Figures 10 and 11, it is apparent that positive values of factors 3 and 4 were associated with greater magnitudes of tongue movement than were negative values. Here a connection between lip kinematics and tongue gestures during vowels can be seen. The vowels which can be characterized as more stiff were produced with less tongue movement than those which would be characterized as less stiff. /ɔ/ had the type of movement captured by factor 4, /æ/ and /o^{U}/ had the type of movement captured by factor 3, and /e^{I}/ had both types of movement. The generalization is this: if there was an appreciable component of tongue movement in the vowel's production, the lip movement was less stiff. The relationship between less stiff opening versus closing movements and this generalization is not obvious. There is no simple relationship like: appreciable movement during the opening phase was associated with a less stiff lip opening movement. This issue must be studied more carefully in another experiment.

Finally note that the within-vowel movement patterns resemble the between-vowel differences in tongue position. When factor 2 was positive (Figure 9, top panel) or when factor 3 was positive (Figure 10, top panel) the tongue dorsum moved up and back during the vowel. This was also the type of difference in tongue position between vowels which had negative values of factor 2 and vowels which had positive values of factor 2. When factor 4 was positive (Figure 11, top panel), the tongue dorsum moved up and forward. This pattern resembles the between-vowel difference in tongue positions found for factor 1 (Figure 1, top versus bottom panel). These observations suggest that both within and between-vowel movements may be driven by physiological constraints such as muscle lines of action. The extent to which this sort of constraint may limit the set of possible sounds in language is an interesting topic which deserves further research.

5. Conclusion

The act of producing a word involves accessing a memory of how to say the word (a long-term representation) and applying that memory to the vocal organs. The hypothesis implied by saying that a long-term memory representation is 'applied' to the vocal tract is that the representation is composed of a system of constraints (Fowler et al., 1980) for the vocal tract not a sequence of fixed motor commands. For instance, the long-term representation of a word may specify that it must start with the glottis closed. If the word is preceded by another word which ends with a closed glottis, the motor commands at the beginning of the second word is different than it is if the first word ended with the glottis open (see MacNeilage & DeClerk, 1969).

The parameters in a system of articulatory constraints are: (1) the movable part or articulator to which the constraint is addressed, (2) the movement goal or target, (3) kinematic properties or dynamics of the movement, and (4) the patterns of coordination among separate movements. Assuming an articulatory constraints view the description of a speech sound involves specifying the parameters in a system of constraints for the sound. Some aspects of the systems of constraints which may be at work in American English vowel production have been identified in the present study.

5.1 *Vowel Articulators*

In most articulatory synthesizers (Mermelstein, 1973; Rubin, Baer & Mermelstein, 1981; Coker, 1976) the location and shape of the tongue is controlled by specifying the location of the center of the tongue which is represented in the sagittal plane as a circle. This strategy was justified by the observation that "the tongue body moves within the mouth as a rather constant-shaped mass" (Coker, 1976, p. 452). This "single-articulator" approach can be contrasted with an approach to describing vowel articulation which is emerging in linguistic theory (Clements, 1985; Sagey, 1986; McCarthy, 1988; Ladefoged & Halle, 1988). In this approach, three different parts of the tongue are described as active articulators for consonants; coronal [t d s z c ɟ], dorsal [k g x ɣ], and radical [ħ ʕ]. This approach to the description of consonants has been extended recently to the description of vowels (Clements, 1990) giving the vowel classifications; coronal [i], dorsal [u], and radical [a].

Is the tongue during vowel production best described as one articulator or several? Interestingly, factor analyses, both the present analysis and Harshman et al.'s (1977) earlier analysis, support both types of description. We find, as would be expected given its incompressibility, that the shape of the tongue is fairly constant. [u] is not produced with a small flap of the tongue raised toward the velum, rather the mass of the tongue moves. So, Coker's (1976) decision to model the tongue as a single mass is warranted by the x-ray data for vowels. Similarly, however, these data also suggest that for different vowels different parts of the tongue form the primary constriction of the vocal tract. For instance, the configuration associated with positive values of factor 1 (top panel, Figure 8) corresponds very well with Clements' (1990) definition of coronal vowels ("produced with a constriction of the tip, blade or front of the tongue" p. 4). Similarly, the pattern associated with positive values of factor 2 corresponds to the definition of dorsal.

Still, a phonetic desciption of vowel articulation in terms of several functional articulators is to be prefered over the single-articulatory approach for several reasons. First, the geometry of the vocal tract dictates that at least three distinct parts of the tongue are naturally inclined to produce closures. The fixed walls of the vocal tract for many speakers have two prominent bends separating three relatively accessible regions (accessible in the sense that the tongue can easily approach the passive wall of the vocal tract). The bends occur between the alveolar ridge and the palate and between the uvula and the pharynx wall. The three accessible regions of the vocal tract then are the alveolar ridge, the hard and soft palates, and the pharynx wall. Obviously, the degree to which this is true for any particular speaker depends on the speaker's palate shape (see Hiki & Itoh, 1986). It is easier to produce a constriction along one of these straight portions than it is to produce a constriction in one of the bends. In fact, for many speakers it may not be possible to produce a constriction in one of the bends (with the tongue body mass) without also producing a constriction at the surrounding areas of the vocal tract. Consequently, even though the tongue can be modelled as a circle moving within the mouth, some parts of the tongue are more capable of producing a constriction than are others. Second, the physiology of the vocal tract dictates that some vowel articulations will be more natural than others. Wood (1979) argued that the cross-linguistic preference for the vowels [i], [a], and [u] can be linked to the physiological/kinematic effects of the the genioglossus (to move the tongue forward and up), styloglossus (to move the tongue back and up) and the hyoglossus (to move the tongue back and down). This model disregards the effects of jaw movement on tongue location which, as was seen above, does seem to play an important role in positioning the tongue. Still, Wood's hypothesis fits nicely with the observation noted above that differences in tongue position between vowels were correlated with typical patterns of movement within vowels. Third, nomograms published by Stevens & House (1955) suggest that there are acoustic quantal regions in the vocal tract. The three regions correspond to the three posited articulators coronal, dorsal, and radical. So, vocal tract anatomy, physiology and acoustics seem to conspire to provide three functionally distinct tongue articulators for vowel production.

5.2 *Vowel Targets*

As discussed in the introduction, the vowels /ɪ, ɛ, ʌ, ʊ/ are phonologically distinct from the other vowels of American English. They do not occur word finally or in open upbeat syllables. One way to describe these phonological phenomena is to consider /ɪ, ɛ, ʌ, ʊ/ as having a single vowel target (V) while the other vowels have two targets (VV). For instance the generalization about word final vowels can then be stated in a two-target analysis: words in English must end in heavy syllables, where heavy syllables are defined as having XVC or XVV structure. The articulatory patterns of American English vowels found in this study are consistent with a two target analysis. Single-target vowels are shorter than two-target vowels and have a shorter vowel nucleus. The data show kinematic organization which reflects the impact of vowel-internal dynamics, and this vowel-internal structure can be described in terms of the number and types of targets in the vowel nucleus. Two target vowels in which the targets differ [e^{I}, o^{U}, ɔ^{ə}, æ^{ə}] have less stiff lip movements than do two target vowels with identical targets [ii, ɑɑ, uu]. In this analysis the lower lip stiffness differences between vowels are the result of two conflicting demands. First, movement toward the second target must be accomplished before the lips close for the final consonant. Second, English prosody demands that two-target vowel not be twice as long as single-target vowels. These conflicting articulatory demands then lead to a reorganization of the lip movement in which its velocity is reduced to allow extra time during the vowel nucleus for the realization of the second target.

5.3 *Vowel Dynamics*

In Browman & Goldstein's (1986) articulatory phonology the vowels [e^{I}, o^{U}, ɔ^{ə}, æ^{ə}] could be implemented by specifying two successive vowel targets during the vowel nucleus. However, it is worth noting that Lehiste & Peterson (1961) found that the vowels in "hide", "how'd", and "hoyd" had two identifiable F2 steady-states, while other long-nucleus vowels did not. This is an interesting observation because the "true" diphthongs involve greater acoustic and articulatory changes during the vowel nucleus than the other long-nucleus vowels, and yet speakers are more likely to produce two steady-states in the 'true' diphthongs than they are in the other long-nucleus vowels. Lehiste & Peterson's (1961) results suggest that speakers can produce vowels with two F2 steady-states (at normal rates of speech, Gay, 1968), even when the targets are far from each other in articulatory space. This suggests that the kinematic reorganization found here for long-nucleus vowels was not a result of having to produce two targets, but rather the result of simply having to produce a movement.

There are several reasons to suspect that movement itself may become phonologized. First, a two-target model of vowel dynamics requires special statements to account for reduction processes for the second (nonsyllabic) target. Several researchers have noted that the formant values at the ends of complex-nucleus vowels are not the same as the formant values at the center of similarly transcribed short-nucleus vowels (Gay, 1968; Gottfried & Miller, 1991). For instance, Gay (1968) found that [ɪ] at the end of [e^{I}] had more variable and different formant values than the [ɪ] of 'hid'. He concluded 'the targets of /ɔ^{I}, a^{I}, a^{U}, e^{I}, o^{U}/ are not necessarily compatible with the vowels used to describe them' (p. 1572). Similarly, in the present data, the offsets of the complex-nucleus vowels did not have the same articulatory positions found for the short-nucleus vowels transcribed with the same symbols. Gay (1968) also found that the formant transitions in complex-nucleus vowels showed much less reduction in duration in fast speech than did vowel steady-states (when steady-states were present at all). Given these observations, then, the phonetics-phonology interface required for the description of vowel targets in complex-nucleus vowels in a two-target model will have to include either separate vowel categories which are only found as non-syllabic vowels in complex-nuclei, or a set of vowel reduction rules which apply only (and obligatorily) to non-syllabic vowels in complex-nuclei. Second, for some voices (women and children primarily) formant movement, or F0 movement is necessary in order for the listener to be able to perceive vowels (Ryalls & Lieberman, 1982). The perceptual problem posed by high pitched voices is that the harmonics of the fundamental are widely spaced in frequency and thus do not specify the vocal tract resonances very well. Ryalls & Leiberman (1982) found that a changing F0 could be used to more accurately specify steady formant values for the listener, but it is obvious that the perceptual problem posed by widely spaced harmonics may also be solved by having the formant values change over time, regardless of the F0 trajectory. Third, Mrayati, Carre & Guerin (1988) discuss the acoustic implications of the incompressibility of the tongue. They suggest that incompressibility results in natural patterns of formant movement because increased constriction in one part of the vocal tract is necessarily accompanied by increased openness in another. So, just as the acoustic resonant properties of the vocal tract cause certain vowels to be more acoustically stable than others (Stevens, 1972, 1989), so also, the physiological properties of the vocal tract appear to cause some formant movement patterns to be more easily produced than others.

Static targets are modeled in a dynamic system as point attractors (Abraham & Shaw, 1989). Current spring/mass models of articulator dynamics (Browman & Goldstein, 1986) fall into this class. The data presented here suggest that vowel dynamics may be more complex than this. One way to model the types of movements which we found here is to string together across time a series of point attractors. Thus, for each vowel we would specify more than one target position. Such a model would definitely give a better fit to the data than would a single-target model, but how many targets is enough and what psychological status do we want to claim for the separate targets? The limit theorem states that as the increment decreases the function becomes better defined. So, as we increase the number of point attractors in our model of a vowel, the series of targets tends to define a function. We may think of the function defined by a series of point attractors as a periodic attractor. The dynamic system defined by a periodic attractor is simpler than the system defined by a series of point attractors because a single function specifies the dynamic properties rather than a system of independent functions. Thus, the system defined by a periodic attractor is more constrained than the system defined by a series of point attractors. Just as a task dynamic model accounts for complex interactions between adjacent gestures using simple control parameters, so a system of periodic attractors may account for the complex movement patterns present in vowel articulation using simple control parameters.

5.4 *Summary*

This study has found that vowel production in northern midwestern English involves several interesting dynamic aspects. The vowels [ɪ, ɛ, ʌ] had shorter vowel durations, shorter vowel nucleus durations, longer closing deceleration durations, greater articulator stiffness, and smaller tongue movement magnitudes during the vowel than did the other vowels. The vowels [ɔᵊ, eᴵ, oᵁ, æᵊ] had the opposite pattern. This leaves just [i, ɑ, u] which had extreme values on the two primary vowel gestures. The results support the view suggested by several acoustic and perceptual studies that dynamic information is important in maintaining distinctions between vowels, because all of the factors of vowel production derived in the canonical discriminant analysis contained aspects of movement as well as general tongue shape, and two of the four factors were associated with movement patterns which divided the vowels into the same classes that resulted from the analysis of lip kinematics. This relationship is particularly interesting because it suggests that the stiffness of consonant releases and closures may be adjusted for the sake of completing a vowel-specific tongue gesture.

Acknowledgements

This paper is dedicated to Ilse Lehiste. Many thanks to Peter Ladefoged and Mona Lindau for sharing their data with me and for many thoughtful discussions throughout the project. Thanks also to Ken DeJong, Bruce Hayes, Pat Keating, Peter Ladefoged and Mona Lindau for their comments on an earlier version of this paper. This research was supported by grants DC00029 and DC00330 from the National Institutes of Health.

References

Abbs, J.H., Nadler, R.D. & Fujimura, O. (1988) X-ray microbeams track the shape of speech, *Soma*, Jan. 1988, 29-34.

Abraham, R.H. & Shaw, C.D. (1989) *Dynamics - The geometry of behavior. Part 1: Periodic behavior.* Santa Cruz, CA: Aerial Press.

Beckman, M.E., Edwards, J. & Fletcher, J. (1991) Prosodic structure and tempo in a sonority model of articulatory dynamics, In *Papers from lab phon II: Segment, gesture, and tone* (G. Docherty & D.R. Ladd, editors). Cambridge: Cambridge Univ. Press.

Bloomfield, L. (1933, 1984) *Language*. Chicago: Univ. of Chicago Press.

Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook,* 3, 219-252.

Chomsky, N. & Halle, M. (1968, 1991) *The sound pattern of English.* Cambridge: MIT Press.

Clements, G.N. (1990) Place of articulation in consonants and vowels: A unified theory. In *L'architecture et la géométrie des représentations phonologiques* (B. Laks & A. Rialland, editors). Paris: Editions du C.N.R.S.

Coker, C.H. (1976) A model of articulatory dynamics and control, *Proc. IEEE*, 64, 452-460.

Di Benedetto, M-G. (1989a) Vowel representation: Some observations on temporal and spectral properties of the first formant frequency, *Journal of the Acoustical Society of America,* 86, 55-66.

Di Benedetto, M-G. (1989b) Frequency and time variations of the first formant frequency: Properties relevant to the perception of vowel height, *Journal of the Acoustical Society of America,* 86, 67-77.

Engstrand, O. (1988) Articulatory correlates of stress and speaking rate in Swedish VCV utterances, *Journal of the Acoustical Society of America,* 83, 1863-1875.

Fowler, C.A., Rubin, P., Remez, R.E., & Turvey, M.T. (1980) Implications for speech production of a general theory of action. In *Language production.* (B. Butterworth, Editor), pp. 373-420. New York: Academic Press.

Fujimura, O., Kiritani, S. & Ishida, H. (1973) Computer controlled radiography for observation of movements of articulatory and other human organs, *Computer Biology Medicine,* 3, 371-384.

Gay, T. (1968) Effect of speaking rate on diphthong formant movements, *Journal of the Acoustical Society of America,* 44, 1570-1575.

Gay, T. (1974) A cinefluorographic study of vowel production, *Journal of Phonetics,* 2, 255-266.

Harshman, R., Ladefoged, P. & Goldstein, L. (1977) Factor analysis of tongue shapes, *Journal of the Acoustical Society of America,* 62, 693-707.

Hiki, S., & Itoh, H. (1986) Influence of palate shape on lingual articulation, *Speech Communication*, 5, 141-158.

Huang, C.B. (1986) The effect of formant trajectory and spectral shape on the tense/lax distinction in American vowels, *IEEE-ICASSP* 86, 893-896, Tokyo, Japan.

Jackson, M.T.T. (1988) Analysis of tongue positions: Language-specific and cross-linguistic models, *Journal of the Acoustical Society of America*, 84, 124-143.

Johnson, K., Ladefoged, P. & Lindau, M. (submitted) Individual differences and universal phonetics.

Kent, R.D. & Moll, K.L. (1972a) Cinefluorographic analyses of selected lingual consonants, *Journal of Speech and Hearing Research,* 15, 453-473.

Kent, R.D. & Moll, K.L. (1972b) Tongue body articulation during vowel and diphthong gestures, *Folia Phoniatrica*, 24, 278-300.

Kent, R.D. & Netsell, R. (1971) Effects of stress contrasts on certain articulatory parameters, *Phonetica*, 24, 23-44.

Kiritani, S., Itoh, K. & Fujimura, O. (1975) Tongue-pellet tracking by a computer-controlled x-ray microbeam system, *Journal of the Acoustical Society of America*, 57 1516-1520.

Kshirsagar, A.M. (1972) *Multivariate analysis.* New York: Marcel Dekker.

Ladefoged, P., DeClerk, J., Lindau, M. & Papçun, G. (1972) An auditory-motor theory of speech production, *UCLA Working Papers in Phonetics,* 22, 48-75.

Ladefoged, P. & Halle, M. (1988) Some major features of the International Phonetic Alphabet, *Language*, 64, 577-582.

Lehiste, I. & Peterson, G.E. (1961) Transitions, glides, and diphthongs, *Journal of the Acoustical Society of America*, 33, 268-277.

Lindau, M. & Ladefoged, P. (1989) Methodological studies using an x-ray microbeam system, *UCLA Working Papers in Phonetics*, 72, 83-90.

Lindau, M. & Ladefoged, P. (1990) Interarticulatory relationships in vowel production, *UCLA Working Papers in Phonetics*, 74, 115-123.

Lindblom, B., Lubker, J. & Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, *Journal of Phonetics*, 7, 147-161.

MacNeilage, P.F. & DeClerk, J.L. (1969) On the motor control of coarticulation in CVC monosyllables, *Journal of the Acoustical Society of America*, 45, 1217-1233.

McCarthy, J.J. (1989) Feature geometry and dependency: A review, *Phonetica*, 43.

Mermelstein, P. (1973) Articulatory model for the study of speech production, *Journal of the Acoustical Society of America,* 53, 1070-1082.

Mrayati, M, Carre, R., & Guerin, B. (1988) Distinctive regions and modes: a new theory of speech production, *Speech Communication*, 7, 257-286.

Nadler, R.D., Abbs, J.H., & Fujimura, O. (1987) Speech movement research using the new x-ray microbeam system. In *Proceedings of the 11th international congress of phonetic Sciences*, Vol. 1, pp. 221-224.

Nearey, T.M. & Assman, P.F. (1986) Modeling the role of inherent spectral change in vowel identification, *Journal of the Acoustical Society of America,* 80, 1297-1308.

Nord, L. (1975) Vowel reduction: Centralization or contextual assimilation? In *Proceedings of the speech and communications seminar, 1974* (G. Fant, Editor). Stockholm: Almqvist & Wiksell.
Parker, E.M. & Diehl, R.L. (1984) Identifying vowels in CVC syllables: Effects of inserting silence and noise, *Perception & Psychophysics*, 36, 369-380.
Perkell, J.S. & Nelson, W.L. (1985) Variability in the production of the vowels /i/ and /a/, *Journal of the Acoustical Society of America*, 77, 123-133.
Peterson, G.E. & Barney, H. (1952) Control methods used in a study of the identification of vowels, *Journal of the Acoustical Society of America,* 24, 175-184.
Rubin, P., Baer, T., & Mermelstein, P. (1981) An articulatory synthesizer for perceptual research, *Journal of the Acoustical Society of America,* 70, 321-328.
Ryalls, J., & Lieberman, P. (1982) Fundamental frequency and vowel perception, *Journal of the Acoustical Society of America*, 72, 1631-1634.
Sagey, E.C. (1986) The representation of features and relations in non-linear phonology. Unpublished MIT Dissertation.
Saltzman, E.L. & Munhall, K.G. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, 1, 333-382.
SAS Institute (1982) *SAS User's Guide: Statistics, 1982 Edition.* SAS Institute, Inc., Cary, North Carolina.
Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data, In *Human communication: A unified view* (E.EE. David and P.B. Denes, Editors), pp. 51-66. New York: McGraw-Hill.
Stevens, K.N. (1989) On the quantal nature of speech, *Journal of Phonetics*, 17, 3-45.
Stevens, K.N. & House, A.S. (1955) Development of a quantitative description of vowel articulation, *Journal of the Acoustical Society of America*, 27, 484-493.
Stevens, K.N., House, A.S., & Paul, A.P. (1966) Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation, *Journal of the Acoustical Society of America,* 40, 123-132.
Strange, W. (1989) Evolving theories of vowel perception, *Journal of the Acoustical Society of America,* 85, 2081-2087.
Strange, W., Jenkins, J.J. & Johnson, T.L. (1983) Dynamic specification of coarticulated vowels, *Journal of the Acoustical Society of America*, 74, 697-705.
Trager, G.L. & Smith, H.L. (1951) *An Outline of English Structure.* Norman, Oklahoma: Battenburg press.
Verbrugge, R.R. & Rakerd, B. (1986) Evidence for talker-independent information for vowels, *Language & Speech,* 29, 39-57.
Wood, S. (1979) A radiographic analysis of constriction location for vowels, *Journal of Phonetics*, 7, 25-43.