Journal of Phonetics (1974) **2**, 255–266

A cinefluorographic study of vowel production

Thomas Gay

University of Connecticut Health Center, Farmington, Connecticut, U.S.A. and Haskins Laboratories, New Haven, Connecticut, U.S.A.

Received 15*th July* 1974

Abstract: The purpose of this experiment was to study the effects of changes in both phonetic context and speaking rate on the movements toward and attainment of target positions for the vowels /i/, /a/ and /u/. Two subjects read lists of nonsense words containing the vowels /i/, /a/ and /u/, in VCV combination with the consonants /p/, /t/ and /k/, at both slow and fast speaking rates. Lateral view X-ray films were recorded along with the acoustic signal. Results showed that during slow speech, the target positions of both /i/ and /u/ remain highly stable across changes in both the preceding and following consonant and vowel. The production of /a/, although not subject to right-to-left effects beyond the following consonant, is sensitive to changes in the consonant, as well as the vowel preceding the consonant. These coarticulation effects, however, are not reflected as such in the acoustical measurements. The production of all three vowels during fast speech is characterized by articulatory undershoot and an upward shift in the frequencies of both the first and second formants. These results are discussed in terms of a target based description of vowels.

Coarticulation has been the subject of considerable interest in recent physiological speech research, yet one that is still little understood. Although it is well documented that variability in the production of a phone exists at all levels of the peripheral production process, there is little data on the exact nature and extent of most coarticulatory phenomena.

One good case in point is vowel production. For example, both MacNeilage & DeClerk (1969) and Harris (1971) have shown that different motor command strategies can be used for a vowel depending upon the phonetic context in which it is placed. However, it is not clear whether these different strategies are reflected in differences in the target positions of the vowel. Indeed, it might be argued that coarticulation at the motor command level simply reflects a strategy to attain a quasi-invariant articulatory target position (MacNeilage, 1970). Unfortunately, however, the available data that bear on this point are somewhat contradictory. For example, the physiological data of Houde (1967), MacNeilage & DeClerk (1969) and Gay *et al.* (1974) suggest that vowel stability is more the rule than the exception, while the X-ray data of Kuehn (1973), suggest a good deal of positional variability for the vowel target (specifically, /a/). Target variability is also evident for faster speech (Gay *et al.*, 1974; Kuehn, 1973) and destressed speech (Lindblom, 1963; Kent & Netsell, 1971).

The purpose of the experiment reported here was to examine more closely a number of aspects of vowel production. The most important of these concerns the nature of a vowel target and whether it can be defined in terms of a three dimensional articulatory coordinate system (MacNeilage, 1970). The experimental approach to this problem was to use cinefluorography to study the effects of changes in both phonetic context and rate of speech on the movements of the tongue and jaw during the production of selected vowels. This was designed to provide a descriptive account of the movements toward and attainment of vowel target positions under the constraints of a variety of linguistic demands known to be sources of articulatory variability. In addition, acoustical analysis was used to determine whether any variability evident at the articulatory level is reflected in the formant structure of the vowel.

Method

Subjects and speech material

Subjects were two adult males, both native speakers of American English. The speech material consisted of the consonants, /p, t, k/ and the vowels /i, a, u/ in a trisyllable nonsense word of the form, /pV_1CV_2pə/ where V_1 and V_2 were all possible combinations of /i, a, u/ and C was either /p/, /t/ or /k/. The 27 utterance type were random ordered into a master list. Each utterance, preceded by the carrier phrase, "It's a ...," was produced at two speaking rates: slow (or normal) and fast. Each rate was based on the subject's own appraisal of comfortable slow and fast rates. A brief practice session preceded the filming session. The subjects were also instructed to produce the first two syllables of the utterance with equal stress, with the final syllable unstressed.

Data recording

Lateral view X-ray films were recorded with a 16 mm cine camera at a speed of 64 fps. The X-ray generator delivered 1 ms pulses to a 9 in image intensifier tube. Two lead pellets (2·5 mm diam.) were attached to the surface of the tongue along the midline. The pellets were located on the dorsum at points approximately 2 and 3 in from the tip. Cyanoacrylate was used as the adhesive. A barium sulfate paste was also used as a contrast medium on the tongue and tantalum was applied along the midline of the nose and lips, to outline those structures. The acoustic signal was recorded on magnetic tape and synchronized with the film record by means of camera generated synch pulses.

Data analysis

The X-ray films were analyzed frame-by-frame, using a Perceptoscope film analyzer. The film was projected life size to a writing surface via an overhead mirror system. Each of the two pellets was tracked in a coordinate system that used fixed landmarks as reference points. These points, along with an outline of the hard palate and upper central incisors, were drawn on a master template. Photocopies of the master were then used as templates for the measurements of each film frame. Measurements were made from the time of /k/ release to the time of closure for the final /p/. Rechecks of a number of measurements revealed a pellet measurement error of no more than 1 mm.

Each of the pellets was tracked in two dimensions: tongue height *vs.* time and tongue backing *vs.* time. Examples of these graphs for Subject FSC are shown in Fig. 1. The relative tongue backing measurements provided little in the way of useful data and will not be presented in this form. As can be seen in Fig. 1, the ballistic patterns for both pellets are essentially the same; the only real difference is a greater amplitude of movement for the anterior pellet during the production of the open vowel, /a/. For no other reason, this pellet will be used to illustrate the data in the Results section.

Jaw movement in the vertical plane was also tracked frame-by-frame. This was done by measuring the vertical distance between the upper and lower central incisors.

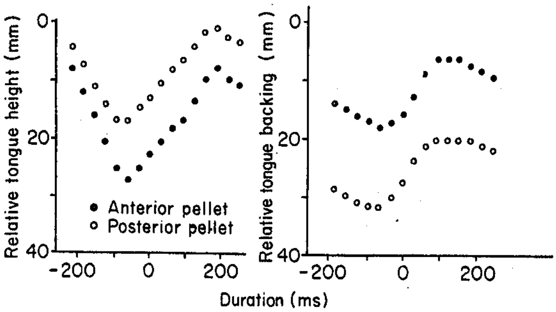


Figure 1 Typical pellet measurements of relative tongue height and relative tongue backing. The utterance is /api/, Subject FSC. 0 on the abscissa = time of closure for /p/.

Besides tracking the dynamics of tongue movement, pellet positions can also be used to construct vowel target positions in the traditional articulatory sense. Figure 2 shows a typical configuration for the vowels, /i, a, u/; the pellet positions appear as the three points in the two dimensional articulatory triangle.

Wide band spectrograms of all the utterances produced during the X-ray run were made from the accompanying magnetic tape recording. Duration measurements were also made from the spectrograms. Durations from the time of /k/ release to the time of /p/ closure averaged 560 ms and 390 ms for Subject FSC and 510 ms and 370 ms for Subject TG, for the normal and fast speaking rates, respectively.

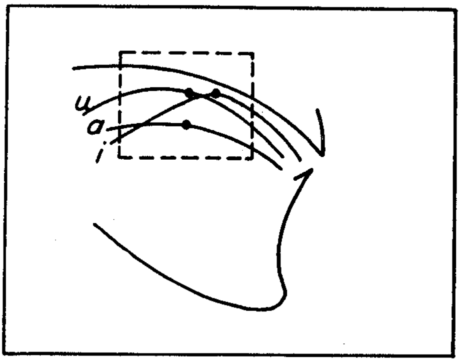


Figure 2 Typical pellet positions at the target positions for /i/, /a/ and /u/.

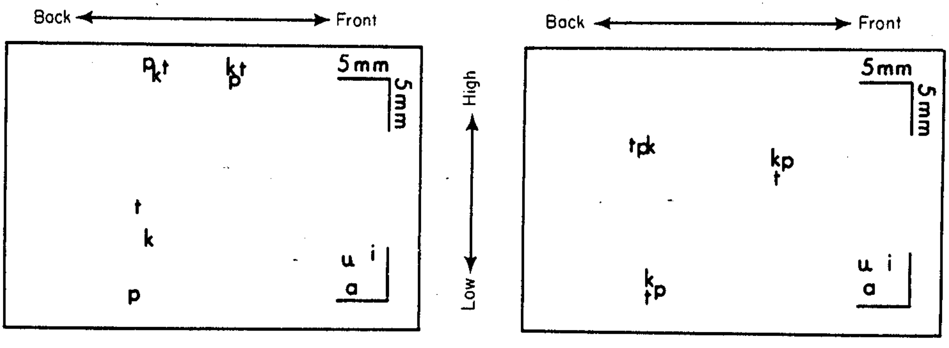
Results

This section will be divided into three parts: the effect of phonetic context on vowel target position, the effect of speaking rate on vowel target position, and the acoustical consequences of these articulatory effects.

Phonetic context effects

In this section the effect of the intervocalic consonant on the target postions of the first and second vowels, the effect of the second vowel on the target position of the first vowel, and the effect of the first vowel on the target position of the second vowel, will be described.

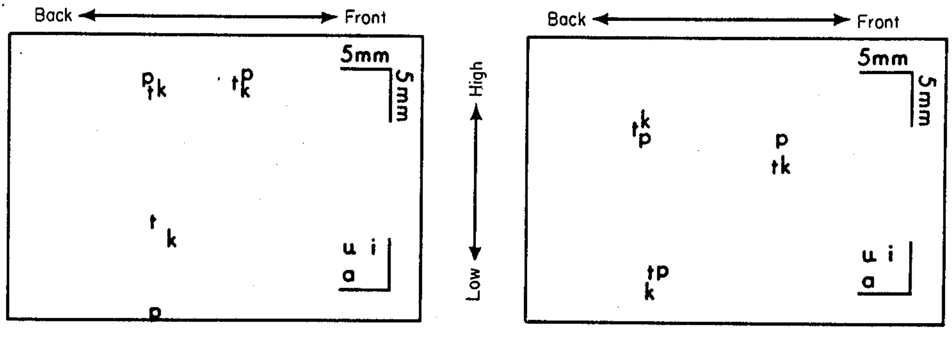
Figures 3 and 4 summarize the effect of the intervocalic consonant on the target positions of the first and second vowels. These figures show the relative positions in two dimensions



Subject FSC

Subject TG

Figure 3 Effect of the consonant on the target positions of the first vowel. The second vowel is /a/. (/ipa, ita, ika, apa, etc./)



Subject FSC

Subject TG

Figure 4 Effect of the consonant on the target positions of the second vowel. The first vowel is /a/. (/api, ati, aki, apa, etc./)

of the anterior pellet at the vowel target (point of farthest articulator displacement). For both subjects, the target positions for /i/ and /u/ in both pre- and post-consonantal positions are quite stable across changes in the consonant. Generally speaking, target variability for /i/ and /u/ rarely exceeded 2 mm, and never exceeded 3 mm. For /a/, however, individual differences appear. While the positions for Subject TG remain stable, the targets for Subject FSC show a rather strong consonant effect in primarily the height dimension. These differences occur for both the first and second vowels and span a distance of almost 8 mm. Displacement for both the first and second vowels is least when the consonant is /t/ and greatest when the consonant is /p/.

The reason for these differences becomes apparent in the movement tracking measurements. Figure 5 shows the measurements for tongue height (anterior pellet) and jaw opening for the entire VCV utterance, for both subjects. While the data for Subject TG show essentially identical movement patterns and target positions throughout the utterance, the data for Subject FSC show variability for all three phonetic segments. This variability appears in both the displacement and velocity components of the curves. The tongue not only extends farther for the vowel, but also moves more quickly (steeper slope) toward its target when the consonant is /p/.

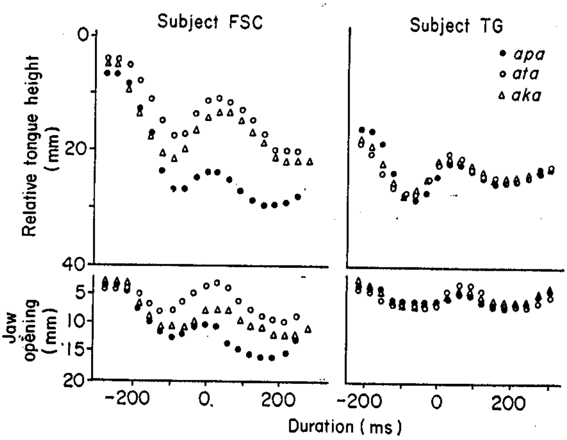


Figure 5

Effect of the consonant on tongue height and jaw opening for both the first and second vowels. 0 on the abscissa = time of closure for the consonant.

Apparently, displacement for the vowel is greatest when the consonant is /p/ because the tongue and jaw are least involved in the production of this consonant. Both /t/ and /k/, on the other hand, are characterized by greater degrees of jaw closure; this probably acts to constrain the degree of opening for the adjacent vowels. Although the displacement differences for the tongue cannot be accounted for entirely by differences in jaw opening, the correlation between the two measures is obviously quite high. This is evidenced by the fact that the curves for the tongue body closely shadow those for the jaw.

In addition to affecting the displacement of its neighboring vowels, the consonant also conditions the timing of the movement toward the vowel. Figure 6 shows the tongue height measurements for the utterances, /ipa/, /ita/ and /ika/. For both subjects, movement toward the second vowel occurs earliest when the consonant is /p/. This effect occurs whenever the tongue moves from /i/ or /u/ to /a/.[1] Again, these differences are apparently related to the independence of the tongue during the articulation of /p/.

Again, too, individual differences in displacement appear for /a/. While tongue displacement for the vowel remains stable in all consonant contexts for Subject TG, con-

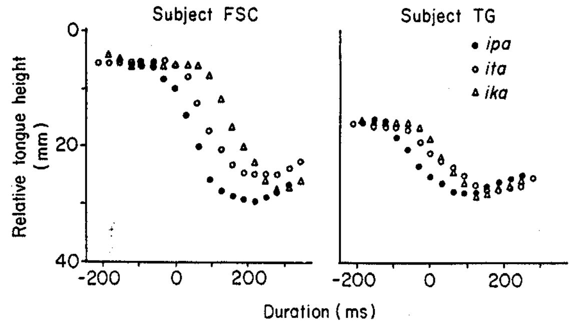


Figure 6

Effect of the consonant on the timing of tongue movements toward the second vowel. 0 on the abscissa = time of closure for the consonant.

---

[1] This effect was also observed in the front-back dimension when the tongue moved from /i/ to /u/, and *vice versa*.

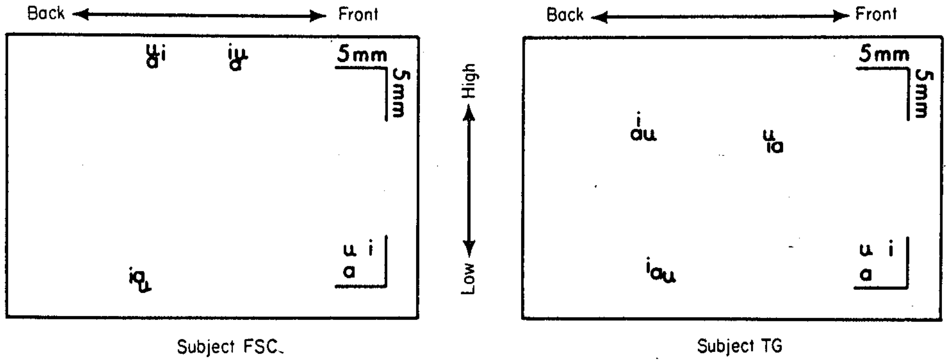


Figure 7 Effect of the second vowel on the target positions of the first vowel. The consonant is /p/. (/ipi, ipa, ipu, api, etc./)

sonant effects are evident for Subject FSC (greater displacement for the second vowel when preceded by /p/, /k/, /t/, in that order). Interestingly, these differences appear to be due solely to differences in the *timing* of the movement from the consonant; in contrast, the effects described earlier were apparently caused by differences in the degree of *displacement* for the consonant.

Figure 7 summarizes the effect of the second vowel on the target positions of the first vowel. This figure shows the target positions of the first vowel as a function of different second vowels in the /p/ consonant context. For both subjects, the target positions of all three first vowels are stable across changes in the second vowel (again, generally within a range of 2 mm). This stability is also evident when the consonant is /t/ and /k/. Apparently, right-to-left effects do not extend across the consonant to the preceding vowel.

Although the first vowel in the VCV utterance is not sensitive to any right-to-left effects beyond the consonant, the second vowel is subject to some left-to-right, or carryover vowel effects; these effects, however, are fairly complicated and linked to the consonant.

When the consonant is /p/ the first vowel has no real effect on the target position of the second vowel (Fig. 8). All three vowels maintain positional stability. However, when the intervocalic consonant is either /t/ or /k/, left-to-right effects appear. Although the targets for both /i/ and /u/ (in the second vowel position) remain stable, the first vowel exerts a strong effect on the target position of /a/, this time for both subjects. These effects are

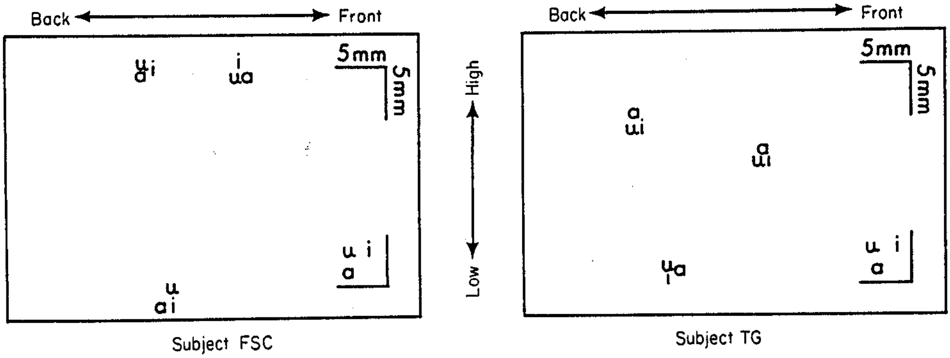


Figure 8 Effect of the first vowel on the target positions of the second vowel, for the consonant /p/. (/ipi, api, upi, ipa, etc./)

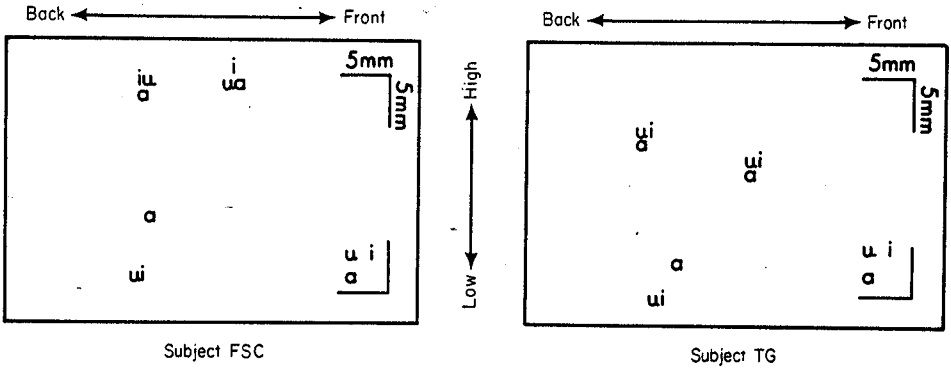


Figure 9 Effect of the first vowel on the target positions of the second vowel, for the consonant /t/. (/iti, ati, uti, ita, etc./)

illustrated in Fig. 9 (two dimension measurements) and Fig. 10 (tongue height *vs.* time measurements) for the /t/ consonant environment. These figures show less opening for /a/ when the first vowel is /a/ than when the first vowel is either /i/ or /u/.

At first glance these effects are quite surprising. It would seem more likely, at least intuitively, that greater degrees of opening for the second vowel would be caused by a more open first vowel. However, closer inspection of Fig. 10 can explain these effects. At the time of closure for the consonant (0 on the abscissa), both the tongue body and jaw are in approximately the same position for each of the three first vowels. Up until this point in time, however, the tongue is *closing* toward this position from /a/, whereas it is *opening* toward this position from both /i/ and /u/. Thus, the tongue is moving in different directions at this point, and, in effect, has a head start towards the second vowel when the first vowel is close.

To determine the extent to which jaw opening controls tongue height for an open vowel, the tongue measurements were plotted against the jaw measurements (tongue–jaw) to obtain the net movement curves for the tongue, i.e. independent from the jaw. These data are shown, for both subjects, in Fig. 11. Since the three vowels in this figure are characterized by almost equal degrees of displacement at and near the target, it is

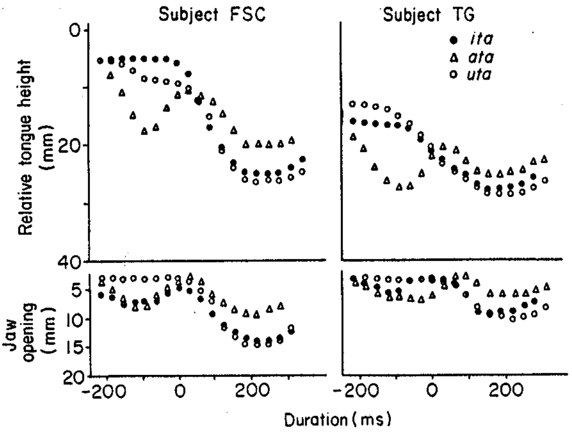


Figure 10 Effect of the first vowel on tongue height and jaw opening for the second vowel. 0 on the abscissa = time of closure for the consonant.

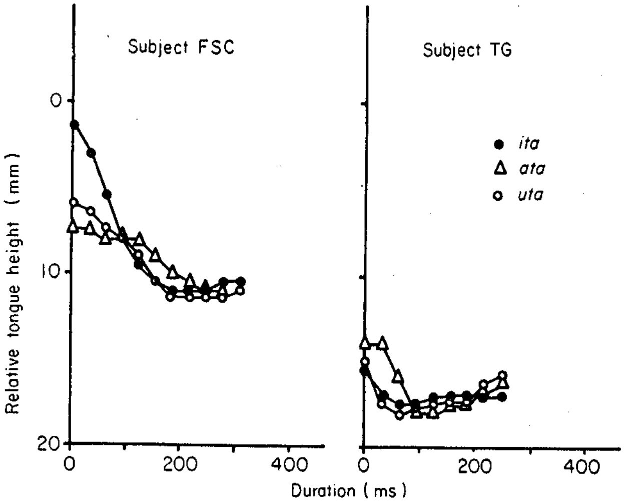


Figure 11 Net tongue movement for /a/. Each data point represents the relative position of the jaw subtracted from the relative position (height) of the tongue. 0 on the abscissa = time of closure for the consonant.

apparent that the differences in opening for the vowel are controlled by the jaw (Lindblom & Sundberg, 1971).

The jaw opening data are interesting from another point of view: /u/ is the only vowel characterized by a closed jaw position. Both /a/ and /i/ are produced with a more open jaw, /a/ for obvious reasons, and /i/, probably to make room for the bunching of the tongue. Although the degree of jaw opening for /i/ shown in this figure approaches that for /a/, in most cases, jaw opening for /i/ is somewhat less than this, usually one-half to, at most, two-thirds that for /a/.

The results of this section can be summarized as follows. The vowel targets of both /i/ and /u/ are highly stable across changes in either the consonant or the vowel. The targets for /a/ are more variable, especially for one subject. Target position variability, when it does appear, is conditioned by both the consonant (left-to-right and right-to-left effects) and the first vowel (left-to-right effects). Right-to-left effects of the second vowel on the first vowel were virtually non-existent.

Speaking rate effects

An increase in speaking rate generally resulted in a decrease in articulatory displacement for the vowel. Although target undershoot was usually present in the speech of both speakers, a number of instances were observed where an increase in speaking rate had no appreciable effect on the displacement of the vowel. These occurrences were more frequent for Subject TG. Undershoot occurred both more often and to a greater degree for /a/, and averaged 3–5 mm for Subject FSC and 1–3 mm for Subject TG.

The context effects that appeared at the slow speaking rate were generally absent at the fast speaking rate. This is probably because jaw movement was more restricted during fast speech; thus, the jaw dependent contextual effects tended to disappear.

Perhaps the most interesting fast speech effect occurred for the /apa/ sequence (illustrated in Fig. 12). For both subjects, tongue movement from the initial /k/ to the second vowel occurs as a single articulatory gesture through both the first vowel and the con-

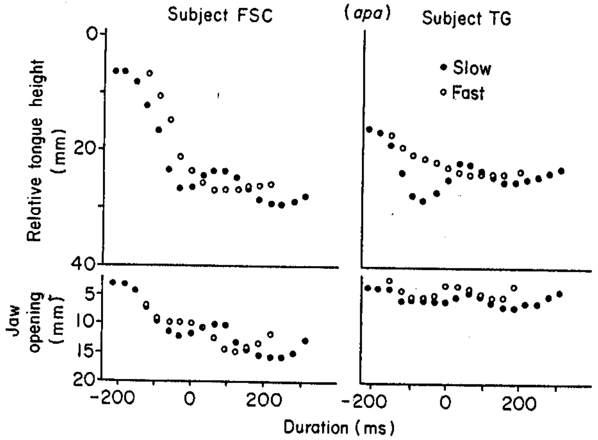


Figure 12 Effect of speaking rate on tongue height and jaw opening for /apa/. 0 on the abscissa = time of closure for the consonant.

sonant; there is no articulatory target evident for the first vowel! In this instance also, the tongue moves somewhat independently from the jaw, with the jaw either holding steady (Subject FSC) or closing (Subject TG) for the intervocalic consonant while the tongue continues moving downward for the second vowel.

Acoustical consequences

Wide band spectrograms were made of all the utterances produced by both speakers. However, because of the noise produced by both the X-ray generator and cine camera, a number of tokens (/u/, in particular) could not be analyzed. In addition, the noise also obscured the first formant frequencies of both /i/ and /u/ to a point where most of these measures must be considered dubious. Nonetheless, the acoustical measures still provided adequate information to shed some light on the two most important questions at issue: (1) whether the context dependent articulatory variability resulted in corresponding acoustic variability, and (2) whether the undershoot effects evident during fast speech reduced the acoustic vowel triangle towards the neutral schwa.

Although the spectrograms showed the presence of considerable acoustic variability, this variability could not be attributed to any of the coarticulation effects described earlier; acoustic variability occurred almost at random. In fact, the only consistent acoustic effect occurred for Subject TG where the first and second formant frequencies of /a/ showed a consonant effect. First and second formant frequencies for /a/ averaged 775 Hz and 1300 Hz when the intervocalic consonant was /p/ and 825 Hz and 1425 Hz when the intervocalic consonant was either /t/ or /k/. It will be remembered, however, that the coarticulation effects of the consonant occurred only for Subject FSC and not Subject TG! The articulatory targets of Subject TG were stable for these utterances. The ranges for first and second formant frequencies across all consonants are shown in Table I.

The effect of an increase in speaking rate on the formant frequencies of all three vowels is shown, for both subjects, in Fig. 13. These graphs show the F1–F2 coordinate positions for all occurrences (where measurable) of /i, a, u/, at both the slow (s) and fast (f) speaking rates.

For both subjects, an increase in speaking rate is accompanied by an increase in the frequency levels of both the first and second formants. The increases are generally greater

Table 1 **Ranges of formant frequencies for all occurrences of /i/, /a/ and /u/ during the slow speaking rate condition. Values are rounded to the nearest 25 Hz**

	Subject FSC		Subject TG	
	F1	F2	F1	F2
/i/	350–450	1850–2125	475–575	2150–2375
/a/	700–850	1150–1375	750–850	1275–1475
/u/	450–525	875–1025	575–625	950–1075

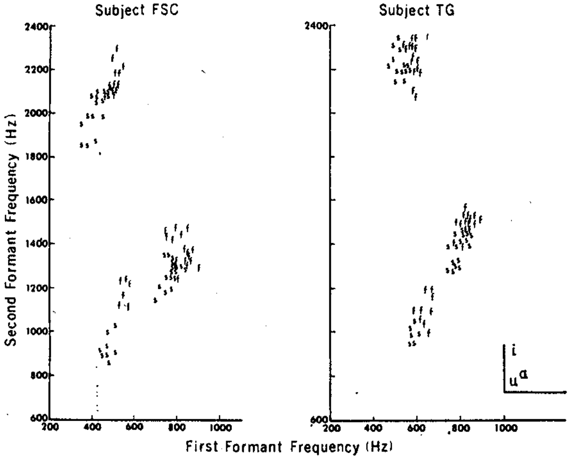


Figure 13 First and second formant frequency measurements for all occurrences of /i/, /a/ and /u/, for the slow (s) and fast (f) speaking rates. All measurements are to the nearest 25 Hz.

for Subject FSC than for Subject TG. The formant frequency measurements for both subjects show the same range of variation during fast speech as during slow speech. Some overlap of coordinate positions is also evident for the two speaking rates. The increases in formant frequencies for /i/ and /u/ might be explained by the more open vocal tract observed for these vowels during fast speech (articulatory undershoot for /i/ and /u/ results in a greater degree of openness). This explanation, however, could not apply to the formant frequency shift observed for /a/. The most important aspect of these measurements, however, is that the acoustic triangle is not reduced towards the neutral schwa during fast speech, i.e. articulatory undershoot during fast speech does not produce the same acoustic result as articulatory undershoot during destressed speech (Lindblom, 1963).[2] These different acoustic effects are probably related to differences in the magnitude of articulatory undershoot characterizing the vowel during fast and destressed speech.

Discussion

The major findings of this experiment can be summarized as follows. During slow speech, the target positions of both /i/ and /u/ remain relatively stable across changes in both the preceding and following consonant and vowel. The production of /a/, although not

---

[2] The notion of vowel neutralization during fast speech does not hold up even if the somewhat dubious measures of F1 for /i/ and /u/ are discounted: the upward shift of F2 for /i/ precludes this possibility.

subject to right-to-left effects beyond the following consonant, is sensitive to changes in the consonant, as well as the vowel preceding the consonant. These coarticulation effects, however, are not reflected, as such, in the acoustical measurements. The production of /i/, /a/ and /u/, during fast speech is characterized by articulatory undershoot and an upward shift in the frequencies of the first and second formants.

These results tend to be somewhat perplexing; on the one hand, there is a strong tendency for target position stability for /i/ and /u/, and on the other, almost as strong a tendency for variability for /a/. The extent of these differences in variability is considerable: maximally, 3 mm for /i/ and /u/, and 8 mm for /a/. Yet acoustic variability for /a/ is no greater than that for /i/ and /u/. This seems to indicate that the acoustical properties of /i/ and /u/ are more sensitive to articulatory variability than those for /a/, at least for the parameters measured in this study. These data somewhat contradict Stevens (1972) quantal theory of vowels: acoustic variability, itself, exists to a considerable degree. On the other hand, though, Stevens' view that opening for an open vowel, for example, can be perturbed considerably without any change occurring in the acoustic output can be accommodated by the present data if the acoustic variability observed were the result of an articulatory perturbation not measured in this experiment (pharyngeal cavity size in particular.)

Perhaps it is the relative acoustic insensitivity to articulatory variability that allows the speech production mechanism a certain degree of latitude in the production of /a/; on the other hand, the effects might be purely inertial. While /i/ and /u/ targets are attained primarily by movements of the tongue, opening for /a/ is controlled by the jaw. It is conceivable that because of its greater mass and the nature of its suspension system, the jaw cannot be moved about with the same degree of accuracy as the tongue.

An increase in speaking rate is also accompanied by articulatory variability; undershoot for the vowel target is more the rule than the exception. However, reduction towards schwa is not evident in the acoustic measures. This, of course, is not necessarily an unexpected result. If vowels produced during faster speech were neutralized, fast speech would be characterized by unintelligible strings of consonants and schwas. Even though undershoot for the vowel is evident for both fast and destressed speech, it is obvious that the two features are controlled by two different strategies. The differences in the acoustic effect are more than likely due to the magnitude of articulatory differences.

Unlike stress effects, speaking rate effects cannot be attributed solely to articulatory sluggishness. The data of both Gay *et al.* (1974) and Gay & Ushijima (1974) show (for the same two subjects used in this experiment) that vowels produced during fast speech are characterized by a *decrease* in the activity level of the muscle; in other words, undershoot is *programmed* in to the gesture. This means, in effect, that a gesture towards a vowel is not directed toward one specific target position. The gesture can be modified to some degree without the loss of whatever perceptually significant acoustic feature limits the vowel. Although the acoustic data during fast speech show an upward shift in both the first and second formants for both subjects, it is not known whether these shifts simply occur within the field for each individual vowel or represent a generalized upward shift of the entire triangle.

Because variability is built into the production of a phone at a level higher than the peripheral speech mechanism, a vowel target cannot be internalized, much less operationally defined, as an invariant event. Nonetheless, MacNeilage's (1970) three dimensional coordinate system still seems to be the best basis for describing a vowel. However, such a specification would have to be expanded to include a spatial field, the boundaries of which are defined by the acoustic limits of the vowel.

Although the data of this study easily fit into a field specified vowel system, the entire schema is by no means complete. First, this experiment studied only the tongue–jaw system; the entire pharyngeal cavity remains unspecified. Secondly, the point vowels, although delimiting, and perhaps normalizing the vowel space for a given individual, can by no means serve to specify all the vowels of a language (Lieberman, 1974). Indeed, before a general model of vowel production can be posited, the intermediate vowels must likewise be specified.

In summary, this experiment produced two major findings. First, articulatory variability in terms of vowel target position exists, but not to a degree where it can be correlated to the acoustic variability that also exists. Secondly, articulatory variability also occurs with an increase in speaking rate; speaking rate effects, however, unlike stress effects, are accompanied by an upward shift in formant frequencies.

The author wishes to thank Dr J. Daniel Subtelny, Head, Department of Orthodontics, Eastman Dental Center, Rochester, New York, for the use of his cinefluorographic facilities. This research was supported, in part, by grants from the National Institute of Neurological Diseases and Stroke (NS-10424), and the National Institute of Dental Research (DE-01774).

References

Gay, T., Ushijima, T., Hirose, H. & Cooper, F. S. (1974). Effect of speaking rate of labial consonant-vowel articulation. *Journal of Phonetics* **2**, 47–63.

Gay, T. & Ushijima, T. (1974). Effect of speaking rate on stop consonant-vowel articulation. *Proceedings of the Speech Communication Seminar, Stockholm, 1974*. Uppsala: Almqvist and Wiksell, 205–209.

Harris, K. S. (1971). Action of the extrinsic musculature in the control of tongue position: A preliminary report. *Haskins Laboratories Status Report on Speech Research* **SR-25/26**, 87–96.

Houde, R. A. (1967). A study of tongue body motion during selected speech sounds. Ph.D. Thesis, University of Michigan.

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

Kuehn, D. P. (1973). A cinefluorographic investigation of articulatory velocities. Ph.D. Thesis, University of Iowa.

Lieberman, P. (1974). On the origins of language: An introduction to the evolution of human speech. *Macmillan Series on Physical Anthropology*. New York: Macmillan and Co., in press.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* **34**, 1773–1781.

Lindblom, B. & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movements. *Journal of the Acoustical Society of America* **50**, 1166–1179.

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.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review* **77**, 182–196.

Stevens, K. N. (1972). The Quantal nature of speech. In (E. E. David & P. B. Denes, Eds), *Human Communication: A Unified View*. New York: McGraw-Hill.