Articulatory Units: Segments or Syllables?*

Thomas Gay+

ABSTRACT

This paper reviews present models of phoneme- and syllable-based models of articulatory programming and the physiological studies of speech production relevant to them. It then goes on to describe the results of some more recent research that is used to refine earlier formulations.

INTRODUCTION

The question of whether the motor input to the speech string is organized in terms of phoneme-size or syllable-size units has been an often studied, yet unresolved, issue in physiological speech research for a number of years. From an experimental point of view, the major obstacle to a solution is the nature of the speech signal that is available for observation. The speech string that presumably enters the articulatory mechanism as a set of discrete phonological units, emerges at the phonetic level as a continuously varying, highly encoded stream. The effect of this encoding can be observed in the production of a given phone in the form of a temporal spreading of its features to, or coarticulation with, adjacent phones. More importantly, these coarticulation effects are quite extensive, spreading beyond both segmental and syllabic boundaries, further obscuring the identity of the basic unit. Nonetheless, formal models of speech programming exist and a body of physiological data is relevant to them. This paper will review these models and a number of experimental findings related to them. It will then go on to describe the results of more recent research that will be used to refine the earlier formulations.


+Also University of Connecticut Health Center, Farmington.

Acknowledgment: The comments and suggestions of, and assistance in collecting additional EMG data from Professor Björn Lindblom and Dr. James Lubker, Department of Phonetics, Institute of Linguistics, Stockholm University, are gratefully acknowledged. This research was supported by grants from the National Institute of Neurological and Communicative Disorders and Stroke (NS-10424), and the National Science Foundation (BNS-7616954).

[HASKINS LABORATORIES: Status Report on Speech Research SR-54 (1978)]
Existing Models

Coarticulation is usually defined as allophonic variation of a phone due to changes in its phonetic environment. These variations, which spread bidirectionally from right-to-left and left-to-right, arise from two different sources. Anticipatory (right-to-left) coarticulation effects are essentially timing effects: for a given segment, movements toward some parts of a feature target begin before others. Carryover (left-to-right) coarticulation effects, on the other hand, are usually considered to be mechano-inertia effects, and exist in the form of variability in target (or target feature) positions as a function of different preceding contexts. Both anticipatory and carryover coarticulation effects appear regularly in speech, and an explanation of both is needed for a general theory of speech production. However, because anticipatory effects originate at an early planning stage of the process, while carryover effects appear after the fact, so to speak, an explanation of anticipatory coarticulation is more central to the question of programming units.

In considering the nature of the articulatory programming unit underlying complex syllable constructions in Russian, Kozhevnikov and Chistovich (1965) proposed the general concept of the articulatory syllable. Using an electro-resistance measuring technique, Kozhevnikov and Chistovich tested their hypothesis by studying the onset of lip rounding (as reflected by lip protrusion) for the vowel /u/ placed in a number of CV, CCV, and CCCV contexts. Each of the segments in the consonant string was unmarked for labiality. Their data showed that the onset of rounding for the vowel usually began during the production of the first consonant in the string even if a syllabic boundary appeared within the string. Since the position of the syllabic boundary was irrelevant to the timing of the protruding gesture, Kozhevnikov and Chistovich advanced the notion that the basic articulatory programming unit in speech was of a CV form with C corresponding to any number of consonants ($C_1 - C_n$), and V corresponding to a vowel. Kozhevnikov and Chistovich further suggested that the motor instructions for each segment within the syllable would be sent out simultaneously, unless the syllable contained competing articulations, in which case the instructions would be issued in sequence.

A second formal model of articulatory programming was proposed by Henke (1966). Using a computer simulation program based on midsagittal x-ray films of the vocal tract, Henke developed a dynamic model that describes the state of the vocal tract at successive points in time. The input to the model is a discrete phonemic string. Anticipatory features associated with a down-stream segment are added to the state of the model at any moment in time by means of a "look ahead" mechanism. This mechanism scans future segmental inputs and issues commands for the immediate attainment of the feature targets of those segments that would not interfere with the attainment of intervening articulations.

Thus anticipatory coarticulation of lip rounding in a CCCV sequence can be explained by the Kozhevnikov and Chistovich model and by the Henke model as well. In Kozhevnikov and Chistovich's model, the onset of lip rounding is considered as part of the set of simultaneous motor commands issued for the production of the entire articulatory syllable. In the Henke model, since the three consonants are unmarked for rounding, the "look ahead" mechanism
searches for the next segment marked for rounding, and issues commands for the rounding gesture to begin during the unspecified consonant string.

While both the articulatory syllable and phoneme-by-phoneme models can explain most of the anticipatory coarticulation effects observed for lip rounding, most investigators prefer to interpret their results in terms of the Henke (1966) model primarily because of its simplicity and greater explanatory powers. While Henke's model specifies a simple phoneme-by-phoneme input, Kozhevnikov and Chistovich's is built around an unnatural and counterintuitive syllable that bears no simple correspondence to common linguistic or phonetic units.

The Forward Spreading of Anticipatory Coarticulation

Anticipatory coarticulation has been studied primarily in terms of three different articulatory features: lip rounding for a rounded vowel, tongue body movements for a postconsonantal vowel in a VCV sequence, and velar lowering for a nasal consonant, with most of the research questioning how far in advance of the particular segment these anticipatory movements begin.

In extending the observations of Kozhevnikov and Chistovich (1965) to American English, Daniloff and Möll (1968) showed that the onset of lip rounding for the vowel /u/ can begin across as many as four consonant segments ahead of the vowel. In their cinefluorographic experiment, the onset of lip protrusion for /u/ was studied for a number of mono- and disyllabic single and two word utterances embedded in sentence frames. Onset of lip rounding usually began with the first consonant in the string, and was not affected by the position of either syllable or word boundaries that appeared in the string. Similar anticipatory lip rounding effects have been demonstrated by Lubker, McAllister and Carlson (1975) at the EMG level, and Benguerel and Cowan (1974) at the movement level. Lubker, McAllister and Carlson's data showed that the onset of EMG activity for the orbicularis oris muscle (a primary lip rounding muscle) associated with /u/ began with the first consonant in the string, and as early as 600 msec prior to the onset of the vowel. In the Benguerel and Cowan study, the onset of lip protrusion for /u/ in a number of similar utterances for speakers of French likewise usually appeared at the time of the first consonant. Benguerel and Cowan also observed that the lip rounding gesture sometimes began as early as the preceding vowel segment. This finding was used to argue specifically against the Kozhevnikov and Chistovich (1965) model; the observed VC coarticulation was inconsistent with the concept of an open CV syllable. However, Benguerel and Cowan did not specify the actual point in time during the vowel when these movements occurred, leaving open the possibility that the movement began during the vowel-consonant transition portion of the vowel.

Another weakness of the Kozhevnikov and Chistovich model is that it does not predict coarticulation effects for certain consonant features--velar lowering, for example--and is not general enough to explain articulatory coarticulation in simple V or VC sequences, strings where anticipatory coarticulation has been shown to exist. For example, in a spectrographic study of coarticulation in Swedish VCV sequences, Öhman (1965) suggested that the variability observed in transition movements from the first vowel to the intervocalic consonant could be predicted by the formant frequencies of the
second vowel. This led Öhman to conclude that the consonant gesture in a VC VC sequence is simply superimposed on a basic vowel-to-vowel substrate. In other words, anticipatory movements toward the second vowel can begin independently of those toward the consonant. Examples of anticipatory coarticulation of velar lowering have also been reported in the literature. Moll and Daniloff (1971) showed that, in a CVVN sequence, velopharyngeal opening for the final nasal usually began during the production of the first vowel in the sequence, that is, two segments in advance of the nasal consonant. McLean (1973) observed similar patterns of velar lowering. In his data, anticipatory velar opening for a final nasal in a CVVN sequence usually began with the first vowel in the sequence, unless the two vowels were separated by a marked junctural boundary.

The studies reviewed above, among others, suggest that articulatory encoding is a complex phenomenon whose effects can spread across several adjacent segments, ungoverned by simple linguistic or phonetic rules. Most support, either explicitly or implicitly, Henke's (1966) articulatory model on the basis that the look-ahead mechanism central to this model can explain the observed coarticulatory variations. However, in recent studies, both cinefluoro graphic and electromyographic evidence was used to argue against the pervasiveness of these effects in general and the operation of Henke's model in particular (Gay, 1975; Gay, 1977a, 1977b).

The Forward Limits of Anticipatory Coarticulation

The two experiments described below were undertaken for the purpose of studying, in greater detail than had been done before, the coordination of articulatory gestures in VC V sequences, at both the articulatory movement and EMG levels.

In the cinefluorographic experiment (Gay, 1977a), conventional high speed (60 fps) lateral view x-ray films were obtained from two subjects who produced various VC V utterances that contained the vowels /i,a,u/ and the consonants /p,t,k/ in all possible combinations. Articulatory movements were tracked by recording the positions, frame-by-frame, of 2.5 mm diameter lead pellets that had been attached to the upper and lower lips, jaw, and several locations along the surface of the tongue, relative to a reference pellet attached at the embrasure of the upper central incisors. This experiment was designed for the specific purpose of exploring the question of whether, in a VC V sequence, an intervening consonant constrains the movements of the articulators, in particular the tongue body and lips, from one vowel to the other; in other words, is the movement from one vowel to another in the form of a simple substrate (Öhman's hypothesis) or is it somehow locked to the consonant, and is the lip rounding gesture for the postvocalic rounded vowel likewise constrained by the intervocalic consonant?

The dynamic properties of articulatory movements in a VC V sequence are illustrated in Figure 1 for an utterance where the intervocalic consonant is /p/. This figure shows the movement tracks of the tongue body, lips and jaw in the height dimension for the sequence /kipap/ as produced by two different speakers. Each track is graphed from discrete points measured every film frame, that is, at approximately 17-msec intervals. Measurements begin during the closure period of the initial /k/ and end at the time of closure for the
Figure 1: Movement tracks (height) for the utterance /kipap/, two speakers. The vertical lines indicate the times of lip closure and release for the intervocalic /p/.

Figure 2: Movement tracks for the utterance /kitip/.
Figure 3: Averaged genioglossus muscle activity for the utterances /kipíp/ and /kítíp/. The vertical line corresponds to the time of voicing onset for the first vowel, and the arrow indicates the trough during the intervocalic consonant.

Figure 4: Movement tracks for /kútúp/.

Figure 5: Averaged orbicularis oris muscle activity for /kútúp/.
Figure 6: Orbicularis oris muscle activity for /utu/ and /ukstu/.

Figure 7: Schematic representation of coarticulatory field for vowel and consonant features.
final /p/. The 0 on the abscissa corresponds to the time of closure for the intervocalic vowel. This figure illustrates the general finding that an intervocalic consonant affects the timing of the movements of the tongue body from vowel to vowel. The movement of the tongue body from the first vowel to the second vowel does not begin until after closure for the intervocalic consonant is completed. This was found to be a salient feature in the production of all the VCV utterances studied. Consonant constraints on vowel-to-vowel movement were as evident in the front-back dimension as in the height dimension; moreover, the same rules that apply to /p/ also apply when the intervocalic consonant is either /t/ or /k/. The only variability in the timing effect appeared in the delay time between consonant closure and tongue body movement. While the lag was usually of the order of 30 msec, it varied anywhere from 0 - 60 msec, depending on the particular utterance.

Perhaps the best illustration of consonant constraints on the programming of articulatory gestures is a VCV where the first and second vowels of the sequence are the same. Figure 2 shows the movement tracks for the jaw and four tongue pellets during the production of /kitip/ for one of the subjects (FSC). Instead of the tongue maintaining the /i/ target position during the consonant, the tongue blade and both tongue pellets show continuous movement during the entire consonant gesture. The blade and anterior tongue body pellet (Pellet 1) appear to shadow movements of the tip, while the posterior tongue body pellet (Pellet 2) moves in the opposite direction, either in a facilitory gesture or towards a tongue body consonant target.

The effects of the intervocalic consonant on the organization of the vowel gesture in a symmetrical VCV are even more evident at the EMG level (Gay, 1975). The average EMG activity of the genioglossus muscle for the sequences /kipip/ and /kitip/ as produced by one of the subjects (FSC) of the x-ray experiment, is illustrated in Figure 3. The genioglossus muscle, which comprises the bulk of the tongue body, is primarily responsible for the protruding and bunching associated with the vowel /i/. This figure shows three separate peaks associated with the utterance. The first peak corresponds to the initial /k/, while the second and third correspond to the first and second vowels. Of particular interest is the deep trough (arrow) that separates the two vowel peaks. The presence of a trough, which signifies a cessation of muscle activity, suggests that the two vowels, although phonetically identical, are organized as two separate events. If the movement of the tongue body during the production of the consonant as observed in the x-ray data (Figure 2) was the result of secondary articulatory influences, positional constancy would still exist at the EMG level in the form of one broad genioglossus peak across the entire utterance. However, the existence of two distinct EMG peaks separated by a deep trough indicates that each vowel is marked by a separate muscle pulse and that the intervocalic consonant plays an important part in the programming of articulatory movements in a VCV sequence. These data argue against Öhman's (1965) hypothesis that implies that the timing of vowel-to-vowel movement is independent of that of the intervocalic consonant. If Öhman's hypothesis were correct, tongue body movements toward the second vowel would begin at about the time of the onset of closing for the consonant. However, movement toward the second vowel begins much later, up to 60 msec after closure for the consonant has already been completed. In other words, the onset of anticipatory tongue body movements for the second vowel in a VCV sequence seems to be limited by the onset of closure of the preceding
consonant.

In addition to placing constraints on the movements of the tongue body from one vowel to another in a VCV, an intervocalic consonant also affects the onset of lip rounding for a rounded second vowel. These constraints are observable at both the EMG and articulatory movement levels (Gay, 1975; Gay, 1977b). In those cases where a rounded vowel appears in a postconsonantal position, the rounding gesture (as reflected by lip protrusion), like tongue body movement, does not begin until after closure for the intervocalic consonant is completed. This is true even for the most sensitive case, namely, a symmetrical VCV containing the same rounded vowels. Figure 4 shows the movement tracks of lower lip height and lower lip protrusion plotted against the same baseline for the sequence /kutup/. Even in this example, it is evident that the rounding feature of the first vowel is not continuous through the consonant. Rather, what appears to be an additional, although small, closing and protruding gesture is superimposed on the rounding pattern. This can be seen as a perturbation in both the lip height and lip protrusion curves during the time of consonant production.

The discontinuity of lip rounding during consonant production is more obvious at the EMG level. Figure 5 shows the corresponding EMG data for the orbicularis oris muscle during the production of the same utterance. The envelope represents the average of some 16 tokens of the utterance. The 0 on the time scale corresponds to the time of voicing offset of the first vowel. This figure shows a deep trough in the EMG envelope during the time of consonant production. Again, the presence of a trough signifies an interruption of muscle activity corresponding to the time of consonant production. While these findings are similar to most of the experimental findings that showed the onset of the rounding gesture to occur during the production of a preceding consonant, or consonants (Daniloff and Moll, 1968; Lubker, McAllister and Carlson, 1975, among others), the presence of the trough argues against the interpretation that the onset of rounding is controlled by a look-ahead mechanism of the type proposed by Henke (1966). Since the first vowel is marked for rounding while the intervocalic /t/ is unspecified, Henke's model would predict that the rounding feature would be retained during the production of the consonant. At the EMG level, this would be reflected by a single broad envelope from the beginning of the first /u/, through the consonant, to the end of the second /u/. Obviously, however, this does not happen. The two vowels are each marked by a separate and distinct muscle pulse, with the onset of the lip rounding gesture for the second vowel constrained, like movements of the tongue body, by the time of closure of the intervocalic consonant.

Because the onset of lip rounding in a VCV sequence seems to occur considerably closer to the vowel than it does in a CCCV sequence, the question arises whether the two types of sequences are governed by different rules. To answer this question, additional EMG recordings were recently obtained from the orbicularis oris muscle during the production of a number of VCV, VCCV, and VCCCC sequences containing the rounded vowel /u/ in both pre- and post-consonantal positions. An illustration of the findings for two extreme cases /utu/ and /ukstu/, as produced by an English-speaking subject, appear in Figure 6. This figure shows both the raw and integrated EMG signals for the upper lip electrode plotted against the acoustic waveform. The integrated EMG
envelope for /utu/ replicates the earlier finding: a trough separates the two vowel peaks. More interesting, however, is the finding that the patterns for the VCCV and VCCCV sequences, illustrated in this figure by /ukatu/, are also characterized by a trough. The first vowel is marked by a burst of muscle activity that ceases when the consonant appears, but resumes almost immediately once the consonant is underway. The pattern illustrated here typifies the slow increase in rounding activity that is usually associated with the second vowel. These findings, which have also been observed for Swedish-speaking subjects by Lubker\textsuperscript{1} provide a convincing illustration that the field of anticipatory lip rounding for the second vowel in any VC\textsubscript{1} - C\textsubscript{n}V sequence seems to be limited by the forward boundary of the consonant category preceding the vowel.

Thus, both electromyographic and cinefluorographic evidence suggest that the relative timing of articulatory movements toward a vowel in VC\textsubscript{1} - C\textsubscript{n}V sequences is affected by the prevocalic consonant(s), even if the consonantal features are not contradictory to those movements. The consonant affects both the tongue body movements toward, and lip rounding gesture for the postconsonantal vowel, in that anticipatory movements toward the second vowel do not begin earlier than the time of closure for the consonant. These findings argue against the appearance of anticipatory movements across phonetic categories and the operation of a look-ahead mechanism for the control of these features.

Do these findings suggest that we turn from a phoneme-based articulatory model to an articulatory syllable model? Not necessarily; the reasons for rejecting the concept of a CV syllable as the basic unit in articulatory programming are convincing ones. Rather, these findings suggest that if the string is organized on a phoneme-by-phoneme basis, the input must be more tightly controlled than Henke's (1966) model specifies. Since the forward extent of anticipatory coarticulation seems to be limited by the forward boundary of the preceding phonetic category, it is reasonable to suggest, that as a general rule, the size of the anticipatory field can be defined solely in terms of the size of that phonetic category.

An illustration of how this rule would operate in speech is shown schematically in Figure 7. The blocks represent successive vowel and consonant features (solid lines) of a hypothetical CVCCV sequence. The coarticulatory field for each segment (dashed) line extends bidirectionally (anticipatory effects to the left and carryover effects to the right) across the segment boundaries. For vowel features such as lip rounding, anticipatory movements can begin well ahead of the vowel and across both syllable and word boundaries, but not across phonetic category boundaries. These field effects are essentially the same for consonants: anticipatory movements associated with consonant production are limited to the field of the preceding vowel category. The duration of the anticipatory field is related solely to the duration of the phonetic category. If the preceding phonetic category contains several segments, the size of the anticipatory field will be larger and the onset of the anticipatory movements will be earlier than if the

\textsuperscript{1}Lubker, J: personal communication.
preceding phonetic category contains only a single segment.

As expressed above, this formulation is similar to Henke's (1966) model, divested of its temporally unspecified look-ahead mechanism. However, it is based largely on data relevant to a single, somewhat peripheral articulatory feature, namely, lip rounding. Obviously, additional data are needed before these findings can be generalized. In particular, rules for the timing of tongue body and jaw movements in consonant clusters, and the effects of stress and speaking rate on the temporal properties of these segmental gestures must be established before the picture can be completed. However, at this point, it is reasonable to speculate that the motor input to the speech mechanism seems to operate by simple rules on phoneme-sized units and within a specifiable temporal field.

REFERENCES


