Physiological Aspects of Articulatory Behavior*

Katherine S. Harris+
Haskins Laboratories, New Haven

INTRODUCTION

The motor theory of speech perception is a statement that we will find a simpler relationship between the string of phonemes that a listener perceives and the articulation of the speaker than between the acoustic signal the speaker generates and perception.

The inconsistencies in the acoustic signal from speaker to speaker and within the speech of a given speaker have three different types of causes--differences in vocal tract size and shape among different speakers, apparent differences in style of different speakers of the same language, and differences in the production of a given phoneme by the same speaker in different contexts.

In principle, the acoustic differences produced by differences in vocal-tract size are quite well understood through a long tradition of study which led to Fant's Acoustic Theory of Speech Production (1960). The modern acoustic theory of speech production permits us to calculate the output sound when the cross-sectional area of the vocal tract is known, and vice versa. Apparently, listeners make some such calculation in extracting messages from speech, and this must be, in part, what enables little children to imitate the speech of adults, even though the shape changes which occur during growth are quite complicated.

Stylistic differences between speakers are of several sorts, having different origins and consequences. First, there are dialectal differences in such things as vowel systems, and second, there is the large class of idiosyncracies that are lumped as speech defects; we will not be concerned with these two sources of phoneme difference further in this paper, however troublesome it may be in practice to sort them out from those discussed.

Our chief concern here will be to discuss differences between contexts in the production of a given phoneme by a single speaker. Can these allophonic differences be assigned to a single phoneme which is present at some level in the nervous system and is variably executed at lower articulatory levels? Furthermore, are phonemes the smallest units of speech storage, or can they be considered, in turn, to be combinations of still smaller invariant parts?

+Also, City University of New York Graduate Center, New York.
This rather general question can be rephrased as a more specific hypothesis. It has sometimes been assumed that each phoneme is stored in the nervous system as a fixed articulatory position, or target. This target is not always reached, but failures of attainment are a consequence of the way the articulatory apparatus operates. Let us explore the operation, to test the truth of the assumption itself.

There have been three general modes of attack on the problem, when phrased in this way. The first is purely acoustic. Both Lindblom (1963) and Öhman (1966) have used acoustic data to make inferences about the nature of shape differences between allophones. This procedure works because the acoustic theory of speech production will generally allow direct inference from acoustic output to shapes, at least within a single speaker. The only problem is that the relationship between acoustic output and articulator shape is sharply non-linear—that is, very small changes in shape from some positions will produce large changes in acoustic output, while in other positions, large changes in shape will produce only small changes in acoustic output. Indeed, this property of the shape-to-acoustic transform has led Stevens (1969) to propose that the articulatory positions of the phonemes have been chosen at locations where small changes in articulation will cause minimal changes in acoustic effect.

A second way of investigating the problem is to look at the movements of the articulatory organs directly. Some methods, such as the observation of the up-and-down movements of the larynx, have a long history in experimental phonetics. Some, such as cinefluorography, are recently developed. All of these are alike in that they are observations of the last stage in the articulatory process before transformation of the articulatory signal into sound and, hence, are equivalent to the acoustic methods discussed above.

A third general technique, also physiological, is examination of the myographic signals generated by the muscles as they contract. This technique is different from those described above in that the signal observed is related not to the position assumed by the articulator but to the force acting on the articulator to bring it to a certain position. A brief description of the electromyographic signal may perhaps make this point clearer.

The muscles of the body are made up of bundles of fibers, organized into what are called motor units. The fibers in each motor unit contract when they receive an impulse from the single nerve fiber which supplies them. Both the nerve impulse and the muscle contraction are accompanied by electrical potential, but the muscle potential, a relatively large signal, is what is observed in ordinary electromyographic (EMG) recording. The essentials of a recording device are an electrode sensitive to differences in electrical potential between two points, an amplifying device, and some form of recorder, which shows a transformation of the potential difference as a function of time.

The electromyographic record shown in the typical study described below is a record of the output of a large number of motor units. In general, the fluctuations of amplitude as a function of time are related to changes in the strength of muscle contraction, but their relationship is by no means a simple one, partly because a stronger contraction is accompanied both by change in the number of motor units firing and by changes in the frequency with which they fire. Furthermore, the size of the recording electrode and its distance from the active muscle fibers will affect the record obtained.
For a somewhat more detailed discussion of the recording techniques used in speech research, the curious reader can see Harris (in press) or, for a discussion of general electromyographic technique, a general text such as Basmajian (1962). It is customary nowadays to present records after computer processing. However, in spite of the trying problems in arranging for such a result (again, see Harris, in press, or Cooper, 1965), the function of the computer is just to provide simple averaging of repeated utterances.

Returning to the relationship between articulator shape and electromyographic signal, it is clear that the transformation that takes place between them is quite complex. As an example, let us assume a given vowel, [i], is represented by a fixed articulatory position (an assumption which is only approximately true). Since electromyographic signals are related to the force of muscle contraction, a movement to the vowel from a consonant with a similar place of articulation, as in the sequence [ti], will involve a smaller signal than from an open vowel, as in the sequence [ai]. This point is made very clearly by MacNeilage (1970) and will be discussed further below. Another complication arising from the nature of EMG signals is that variation in size will be associated both with speed of articulator movement and with distance. Generally, a larger signal is associated with a faster, as well as a larger, movement.

In spite of these complications, there seem to be three rather compelling reasons for studying the motor patterns of articulation. First, as Cooper (1965) and Liberman et al. (1967) have pointed out, they are one step closer to the speech-generating center in the brain than are articulatory shapes. Second, most articulators are rather inaccessible; in some cases, it may be easier to examine EMG signals than other physiological variables. Third, even when only anatomical data about the muscles are available, they allow some insight into the articulatory shapes that the speaker can generate.

This diversity of uses of muscle study has resulted in a diversity of purpose for studies already performed; some are aimed at simply describing the mechanism for a given action and some at the more general purpose of understanding speech generation. In what follows, we will try to sketch the anatomy of the articulatory muscles and then to discuss what EMG studies, and the other closely related physiological research, have to offer for understanding the organization of speech. For anatomical detail, the interested reader can refer to more general speech textbooks, such as Zemlin (1968) or van Riper and Irwin (1958) or standard anatomical works.

THE ORGANIZATION OF THE SPEECH MUSCULATURE

The speech musculature can be divided into three more or less independent groups—respiratory, laryngeal, and articulatory.

The respiratory muscles act to provide power for the speech mechanism, the laryngeal muscles act to transform the power into acoustic energy, and the articulatory muscles modulate this carrier to produce the specific sounds of speech, though the functions of these three groupings overlap. We will discuss each of the three in turn.

The Respiratory System

In normal breathing, inspiration and expiration occur approximately
equal amounts of time. In speech, the inspiration-expiration ratio is considerably changed, so that the expiration occupies about seven-eights of the total cycle. The mechanisms of breathing seem to set about four seconds as the limit on an expiration, without a necessary pause for inspiration. This duration can be considered as a sort of physiological limit of phrase length.

The lungs can be thought of as a pair of balloons which are inflated for inspiration and deflated for expiration. These balloons are encased in a partly bony, partly muscular cylindrical cage. The ribs run around most of the cylinder, with a double muscular layer, the internal and external intercostal muscles, running between them, while the diaphragm, a large, dome-shaped muscle, forms its bottom. At the top, stems of the two balloons join, the larynx forms a valve at the top of the connecting, inverted Y-shaped tube, the trachea. In inspiration, the size of the lung cavity is increased by the contraction of the diaphragm and the lifting of the rib cage by the contraction of the external intercostals (and other muscles) (Ladefoged, 1967; Draper et al., 1958). When the lung cavity is increased in size, air flows into the lungs. Phonation occurs when the vocal folds are placed over, and set vibrating by, the inspiratory air stream. To some extent, expiration is a passive phenomenon; the lungs are elastic in nature and, consequently, once inflated, will tend to deflate themselves. The functions of the muscles of respiration are organized in an ancillary way around this function. At the beginning of phonation after a deep inspiration, the muscles of inspiration can be used to break the outflow of air. At the end of expiration, the muscles of expiration (the internal intercostals and auxiliary muscles) can be used to prolong the breath-group, by squeezing more air out of the lungs, acting in opposition to the tissue forces which resist the deformation of the chest wall.

The point at which there is a changeover from the use of expiratory muscles to inspiratory muscles depends on the subglottal pressure maintained. At any time during the breath group, the internal intercostals can be used to produce momentary stress peaks (Ladefoged et al., 1958). Generally, subglottal pressure does not remain constant within a breath-group but falls at the end (Lieberman, 1967).

The effects of variation in subglottal pressure are two-fold. First, subglottal pressure affects the intensity of speech. Secondly, it can be shown that, everything else being equal, greater subglottal pressure will produce higher fundamental frequency. The mechanism for this process is described by Lieberman (1967). There are, then, three effects of speech respiration on the organization of speaking. First, the duration of expiration that can be sustained without an inspiration provides a physiological bound on phrase length; second, a respiratory mechanism permits the assignment of heavy stress; and third, an available subglottal mechanism accounts for the terminal pitch fall at the end of sentences.

The last two points have been the source of a good deal of discussion and controversy, which must be further discussed in connection with laryngeal mechanisms. For the present, let us just comment with respect to what we know about subglottal action itself.

First, Ladefoged has shown that, when a heavy stress is placed on one word in a sentence, there will be an accompanying increase in the action of the internal intercostal muscles. This should have the effect of producing
the brief peaking in the subglottal pressure curve noted by Lieberman (1967) and Ohala (1970). There is no argument that this, in turn, will produce a peak in intensity in the acoustic speech output or that intensity rises, in turn, following Fry (1955), are one of the acoustic correlates of perceived stress. However, the subglottal peaks have been demonstrated only for very heavy, or contrastive, stress, which may not be a very central maneuver in ordinary running speech. Furthermore, there has been a general debate as to how large the effects of such peaks in subglottal pressure are on the fundamental frequency contour (Ohala, 1970), which is perhaps a more important correlate of perceived stress than is intensity (Fry, 1955).

The fall in subglottal pressure at the end of ordinary breath-groups is easily accounted for by the passive nature of the expiratory mechanism. If it is not compensated by the action of the laryngeal muscles, it will result in a fall in fundamental frequency, although here, again, there is argument about how much of the observed acoustic effect is subglottal in origin (see Ohala, 1970).

Laryngeal Mechanisms

Detailed description of the larynx is presented in Sawashima's chapter of this volume. Here, we will confine our attention to speech studies of the laryngeal muscles, particularly electromyographic studies. If this discussion is to be intelligible, however, it will be necessary to summarize some of the more general literature on laryngeal functions. The functions of the intrinsic muscles have been clarified by a long series of experiments by van den Berg and Tan (1959), in which they reconstructed the larynx of cadavers, duplicating normal air flow conditions, and modeled the effects of contraction of the various muscles. They describe the intrinsic muscles as having three general functions: tensing the vocal folds (the cricothyroid and vocalis), adducting the folds (the vocalis interarytenoids and the lateral cricoarytenoid muscles), and abducting the vocal folds. In general, the electromyographic studies of the action of these muscles in singing, from Faaborg-Anderson (1957) to the more recent work of Hirano et al. (1969) and Sawashima et al. (in press), have shown that the adductor muscles and the cricothyroid act together when pitch rises.

The action of the extrinsic laryngeal muscles is far less well understood. These are muscles which originate outside the larynx but insert on one of the laryngeal cartilages and hence influence the tension on the vocal folds and, consequently, pitch, indirectly. However, the direction of the influence will depend on interactions between the muscles above and below the larynx. For example, Hirano et al. (1967) and Hirano et al. (1969) have shown that one member of this group of muscles, the sternohyoid, is active at both high and low pitch extremes in singing. Sonninen (1956) has shown that persons who have had the sternohyoid, thyrohyoid, and omohyoid muscles sectioned for medical reasons typically have trouble singing high pitches after surgery. On the other hand, if the thyrohyoid, one of this group of muscles, is stimulated during surgery, pitch lowering may result, depending on the subject's head position.

Turning now to speech studies, the action of the laryngeal muscles has been studied only in a very few circumstances. There seems to be general agreement about what happens when a word is stressed in an English declarative
sentence, resulting in an upward excursion of the pitch contour. As Hirano et al. (1969) have shown, there is, typically, a burst of activity in the cricothyroid, lateral cricoarytenoid, and vocalis muscles, accompanied by a peak in the fundamental frequency contour. Thus, for heavy stress, there is coordinated activity of laryngeal and subglottal systems to produce a peak in fundamental frequency and intensity, although Ohala (1970) believes that the effects of the subglottal system in producing the pitch rise are negligible. All three of these muscles can be shown to be active in the characteristic terminal rise for questions.

Considerable controversy surrounds the issue of pitch-lowering mechanisms in speech. The question can be subdivided into two substantive issues. First, is there an active pitch-lowering mechanism which is used for producing sudden downward pitch excursions in the course of an utterance? Second, is such a maneuver responsible for the fall in pitch at the end of declarative sentences?

English is perhaps not the ideal language for answering the first question, since, although Bolinger (1958) has demonstrated that sudden downward excursions in fundamental frequency are sometimes used to signal stress, this is not the common maneuver. Better examples are provided by Swedish, where a word accent distinction is signaled by variations in pitch contour. A model for Swedish word intonation has been suggested by Öhman (1967a). He suggests that the effects of Swedish grave and acute accents can be derived from a model that has positive sentence-intonation pulses and negative word-intonation pulses, which are differently timed for grave and acute accents and for different Scandinavian dialects. The model itself suggests, although it does not, of course, require, both a mechanism for pitch raising and an active mechanism for pitch lowering.

Two electromyographic investigations of Swedish word accents have been made. In the first, by Öhman et al. (1967), the cricothyroid and vocalis muscles were studied. Based on the results cited above, we would expect both cricothyroid and vocalis activity to be correlated with pitch rises. The most notable result of the study is an indication that there is a period of inhibition of cricothyroid activity which corresponds to the different times of application of the posited word intonation filter. Thus, there appears to be an inhibition of pitch-raising mechanisms that is correlated with pitch falls. No striking results were obtained for the vocalis.

A similar experiment was performed by Gårding et al. (1970), with probes in vocalis, cricothyroid, and sternohyoid muscles. Generally speaking, they find correlation between peaks in fundamental frequency and in the activity of cricothyroid and vocalis muscles. They find that "the sternohyoid activity shows no simple correlation with the pitch value."

Another language where sudden downward pitch excursions have a linguistic function is Tokyo Japanese, where the pitch level drops at the boundary following a vowel with an accent kernel mark. An electromyographic study of the cricothyroid, the lateral cricoarytenoid, and the sternohyoid was performed by Simada and Hirose (1970). In general, they find that there is a sharp fall in cricothyroid activity corresponding to the accent kernel. The activity pattern of the lateral cricoarytenoid is similar, although complicated by the participation of the muscle in voicing gestures. Most of their data do not show a clear correlation of the activity of the sternohyoid with the position of the accent kernel.
The experiments described above appear to indicate a passive pitch-lowering mechanism; that is, when pitch falls, the activity of the cricothyroid and the muscles that provide medial compression decreases. However, the only muscle that might lower pitch by increasing its activity which has thus far been examined, the sternohyoid, does not seem to indicate a clear pattern of correlation with pitch fall. It may be that other muscles are implicated.

The situation with respect to laryngeal adjustment at the termination of sentences is complicated by Lieberman's (1967) suggestion that the fundamental frequency fall is due to the speaker's failure to compensate at the larynx for falling subglottal pressure. This suggestion would be negated by either active or passive laryngeal adjustment. Speakers could either decrease the activity of the cricothyroid and its associated muscles at the end of a sentence or increase the activity of the sternohyoid, or some muscle with a similar activity pattern, at the ends of sentences. Inspection of cricothyroid records from Ohala's (1970) and our own (Harris et al., 1969) work does not reveal a characteristic fall-off in cricothyroid activity at sentence termination. Ohala (1970) has suggested that the sternohyoid has a tendency to be more active at sentence termination, but the picture is complicated, according to the later work of Ohala and Hirose (1970), by the tendency of the sternohyoid to participate in segmental gestures, such as jaw opening for open vowels. Again, other possible active muscles have been suggested for the pitch-lowering function.

In summary, then, the mechanism for raising pitch in speech has been demonstrated several times. When the cricothyroid contracts, pitch rises. Other muscles that contribute to medial compression of the vocal cords also contract, except as their function is complicated by their participation in voicing gestures. Pitch falls when the muscles that raise pitch relax or when subglottal pressure falls. In addition, an active mechanism for pitch lowering has been suggested, though not conclusively demonstrated. A further complication, not discussed here, is that actual pitch contour is influenced by the shape of the upper vocal tract (Flanagan and Landgraf, 1968).

The Upper Articulators

The third great group of articulatory muscles are those that are responsible for generating segmental phonemes. They, in turn, can be divided into a palate group, a tongue group, a group responsible for raising and lowering the jaw (which we will not describe), and those muscles of facial expression that act to mold the lips. For the linguist, it is probably confusing and unnecessary to supply an enormous list of muscle names and functions, cribbed chiefly from anatomy texts, which are not at present fleshed out by many physiological studies of muscle function in speech. However, sketching the anatomy of the oral region gives one a somewhat better idea how muscular organization limits vocal-tract shape. For the anatomically inclined, van Riper and Irwin (1958), some years ago, made some detailed guesses as to the muscle action involved in forming the vowels and consonants of American English. It is hard to see how these guesses could be improved on by anything except positive information.

The Tongue. Let us begin with the tongue. It has two great muscle systems--extrinsic muscles, which connect the tongue to another structure, and intrinsic muscles, whose fibers run entirely within the tongue body.
The tongue can be moved forward by the genioglossus muscle, a fan-shaped muscle whose fibers make up a great part of the core of the tongue. It can be moved up and back by the styloglossus, which runs from the sides and back of the tongue to the styloïd process of the skull just behind the ear. It can be moved down and back by the hyoglossus, which runs from the sides of the tongue to the horns of the hyoid bone, a horseshoe shaped bone which forms an underpinning for the tongue. In addition to direct pull in these three directions, the tongue may be pulled in the same directions by muscles connecting the hyoid bone to other structures.

The intrinsic muscles of the tongue are named for their fiber directions: the transverse and vertical muscles, which probably groove and flatten the tongue, and the inferior and superior longituinals, which probably curve the surface up and down. It is generally believed that these muscle fibers act together to shape the tongue tip.

There has been very little electromyographic work on the tongue. The technical problems involved are quite difficult; in particular, the question of what muscle is being examined when a probe is inserted makes assignment of function very difficult. The recent development of a flexible wire electrode (Hirano and Ohala, 1969) should greatly facilitate work of this type in the future. However, our knowledge of the movements of the tongue and its associated structures has been greatly enlarged by two recent cineradiographic studies by Houde (1967) and Perkell (1969). Most of their discussion is concerned with general models of the articularatory process; two specific hypotheses might perhaps better be described here.

The first is the observation by Perkell that the vowels of American English conform to the tense/lax description proposed originally by Jakobson et al. (1963) and discussed at greater length by Chomsky and Halle (1968). The latter state: "Tense sounds are produced with a deliberate, accurate, maximally distinct gesture that involves considerable muscular effort; non-tense sounds are produced rapidly and somewhat indistinctly" (p. 321). Perkell believes that he observes such differences in vocal tract adjustments of pairs such as /i/ and /I/. He cites the work of MacNeilage and Sholes (1964) on the electromyographic activity of the tongue to provide further support for this notion; they show that tense vowels have generally higher EMG voltage levels than their lax counterparts.

Indeed, the MacNeilage and Sholes data show duration differences between /i/ and /I/ quite clearly, as one might expect. However, if one examines their data, the ranking of total EMG activity over all the vowels does not conform to a tense/lax classification. Furthermore, the MacNeilage and Sholes data are taken from a very restricted set of sampling points and, consequently, cannot be used for inferences about activity in the tongue musculature as a whole.

Perkell suggests also that the extrinsic muscles of the tongue are used for vowel production, while both intrinsic and extrinsic muscles are used for consonants. He points out that the tongue behaves like a semirigid body in the production of vowels: its shape is more or less constant, and it is moved into target position by the extrinsics. The same observation of the constancy of the tongue shape is made by Houde (1967). At present, it is not possible to assign this result to a specific muscle set. Perkell believes
that the intrinsic tongue muscles are particularly implicated in consonant production. It seems reasonable that the intrinsic muscles of the tongue should be responsible for the complicated shaping of the tip in producing apical consonants. It is less easy to see how they could be primarily responsible for /k/ closure. In this case, the extrinsic muscles may well pull the body of the tongue up and back, without changing its shape in detail. Of course, any such hypothesis must wait on the development of better techniques for electromyographic study of the tongue muscles. It would be interesting to evaluate the hypothesis with a wider range of back tongue consonants than Perkell uses.

The Velopharyngeal Closure System. Velopharyngeal closure has probably been more extensively studied by physiological techniques than has any other part of the speech mechanism. The reasons for this have to do with the clinical problems of its repair.

The action of the palate in making velopharyngeal closure has been studied by both X-ray and electromyographic techniques. X-ray studies are summarized by Bjork (1961) and Nylen (1961). Electromyographic studies are summarized and extended in a recent monograph by Fritzell (1969).

Briefly, velopharyngeal closure is accomplished by elevating the soft palate to block the nasal passageway when an oral sound is produced. The chief active agent would appear to be the levator palatini muscle, which makes up the bulk of the soft palate (Fritzell, 1969). Older accounts (Bloomer, 1953) suggest that velopharyngeal closure is accomplished by a kind of "purse-string" action—that is, the sides of the pharyngeal port come in as the palate elevates, while the posterior pharyngeal wall comes forward. More recent X-ray studies suggest that this purse-string action of the posterior pharyngeal wall is not common in normal speakers (Hagerty et al., 1958). Furthermore, one EMG study (Harris and Schvey, 1962) has shown that, while the muscles of the upper pharyngeal wall are quite active in speech, this activity is not well correlated with the action of the palate. However, it can be shown to be a mechanism occasionally used in effecting velopharyngeal closure in persons with insufficient tissue to make the closure with the velum alone. This difference in mechanism is interesting from the point of view of the general theory of phoneme formation, since it suggests that a wide range of individual differences in gesture are possible for the achievement of the same result.

A second interesting question about articulatory dynamics has been raised in discussion of the nature of velopharyngeal opening after closure. The situation is very much like that for pitch lowering. It is not clear whether velar opening is under active muscular control or whether it is accomplished by gravity, in addition to the relaxation of the muscles of velar closure. Fritzell (1969) has some preliminary data which suggest that there is active contraction of muscles which lower the velum when a nasal follows an oral consonant—that is, that there is active velar lowering—but further research is clearly necessary on this point.

Another question that has been studied is how many different degrees of velar activity are necessary in speech—i.e., whether there is a partial reorganization of the oralization gesture which depends on other aspects of the phoneme. Fritzell (1969) presents clear evidence which supports the earlie
work of Lubker (1968) by showing that the extent of palate movement for high vowels is greater than for low vowels—that is, that the oralization gesture is reorganized. We will discuss this result further below.

Facial Muscles. The third great group of muscles involved in articulation are those that shape the lips—also known as the muscles of facial expression, for obvious reasons. Generally speaking, the muscles shaping the mouth can be conceived as falling into two functional groups: the orbicularis oris fibers, which form a heavy band around the mouth opening and act in sphincter fashion to round and purse the lips, and a series of muscle bundles that insert radially into the oris bundle from various directions. These bundles have an action that depends on their insertion—to pull the upper lip up, to pull the lower lip down, and to spread the lips laterally. In general, all the labial consonants of English show an implosion peak of orbicularis oris activity and an explosion peak in the muscles that withdraw the lips.

In general, it has been found that the size of the closure peak for English /p/, /b/, and /m/ appears to be the same regardless of the following vowel or which of the three consonants is being produced (Harris et al., 1965; Fromkin, 1966; Tatham and Morton, 1968). One study (Öhman, 1967c), however, has found a small difference between closure peak sizes for /p/, /b/, and /m/ (using a Swedish speaker). The size of the explosion peak in the muscles that withdraw the lips varies systematically with the following vowel (MacNeilage and deClerk, 1969). All these results will be discussed below in connection with theories about motor organization in speech.

THE ORGANIZATION OF SPEECH

Both MacNeilage (1970) and Öhman (1967c) have concluded that a description of articulatory dynamics should have the phoneme as its basic unit, in spite of the elusiveness of its physical manifestations. As we have seen, there is nothing necessary about this view of the articulatory process. Mattingly's chapter of this volume summarizes some of the problems of this point of view as a scheme for speech synthesis. In this section, we will try to indicate the problems in view of what is now known about physiological articulatory phonetics.

Allophonic Variants

We have already discussed MacNeilage's (1970) suggestion that if the phoneme is to be considered as a basic unit, it must be stored in the nervous system as a positional target rather than as a movement. He suggests that the motor system is controlled by the results of an internal specification of certain spatial targets. This accounts, presumably, for the results of MacNeilage and deClerk (1969), who showed that, in the production of CVC monosyllables, the lingual electromyographic signals for the vowel are regularly conditioned by the preceding consonant. Fromkin (1966) and Öhman (1967c) have shown similar results for the muscles controlling the lips. For example, Fromkin showed that the amount of activity associated with the rounded vowel /u/ is greater in the context /dud/ than in the context /bub/, presumably because in the latter context the lips are closer to the rounding target for the vowel than in the former.

Given that the articulators tend to adopt a fixed position for a given
phoneme, it is still in question whether stress allophones of the phonemes are commonly preserved— that is, whether phonemes are more strongly articulated in positions of heavy stress. The evidence for stress allophones is presented in connection with the section "undershoot," below.

The question of stress comes up in connection with the "feature" argument. Jakobson et al. (1963) proposed that certain consonants and vowels differ, pairwise, from each other in a "tense/lax" dimension. The "tense" member of the pair was supposed to be a more forceful articulation than the "lax." This distinction has already been discussed relative to the vowels. The same distinction, in their system, is meant to be the primary distinction between "voiced" and "voiceless" consonants. In particular, /p/ is meant to be distinguished from /b/ primarily along a tension dimension. Based on this assumption, one would expect that the orbicularis oris contraction for the /p/ closure would be more forceful than for the /b/ closure. This expectation has been examined in four studies (Harris et al., 1965; Fromkin, 1966; Tatham and Morton, 1968; Öhman, 1967c). In the first three, differences between /p/ and /b/ have been found to be insignificant, while a very small peak-contraction size difference was found in the fourth.

Muscular tension differences have been suggested as an explanation for /p/ - /b/ distinctions in quite another way. Chomsky and Halle (1968), followed by Perkell (1969), have suggested that the observed difference in upper pharyngeal tract size is a passive response to the fact that the upper tract muscles are "tenser" and hence hold the tract walls more rigid for the tense consonant. The larger pharynx size for voiced consonants is confirmed by Kent and Moll (1969). They feel, however, that the size adjustment is under active muscular control: the extrinsic muscles act on the larynx to lower it and hence increase the size of the upper vocal tract. The details of this extremely complicated physiological argument are discussed in Lisker and Abramson (1971) and in Lisker's chapter of this volume. The weight of the evidence seems to suggest that mechanisms other than generalized tensing are responsible for the perceived voiced/voiceless distinction.

It is hard for me to imagine the detailed workings of a system in which both stress differences and phonemic differences were maintained by the same general physiological mechanism. My own present view is that the voiced/voiceless distinction is carried by the timing of glottal adjustment, while the peripheral articulation of cognate pairs is the same. This implies that place and voicing features can be spatially segregated, and it also implies general articulatory separation of features.

A preliminary study (Harris et al., 1962) suggested that the oralization gesture might be the same for voiced and voiceless consonants and, thus, similarly preserves the independent organization of features. However, Lubker (1968) has shown that, for vowels, the oralization gesture, that is, the size of the velopharyngeal closure gesture, is not independent of the vowel height. Therefore, we cannot assume a general orthogonal articulatory organization of features.

If we assume that phonemes are stored as constant targets, we must develop ways of specifying and explaining the obvious failure of the articulators to move from one invariant target to the next in running speech. Using Öhman's
terminology, three mechanisms have been proposed to account for allophonic variation—reorganization, undershoot, and coarticulation.

Reorganization

"Reorganization" is a grab-bag term intended to cover cases where allophones have different articulations in a fundamental sense—i.e., there is a context-dependent change in feature specifications. The two examples given by Öhman, devoicing of final voiced consonants in German and Russian and quality alternations of vowels under vowel harmony, are language specific; at present, no general theory of speech production accounts for the particular circumstances in which reorganization should occur. However, the other two mechanisms are presumed to have universal application, and some general statements have been made about them.

Undershoot

Undershoot results when "an incomplete articulatory gesture is interrupted by a neural command that brings about the next gesture of the utterance." The key experiment for the demonstrations of undershoot was performed by Lindblom (1963) on vowel reduction. He had subjects produce CVC syllables in a sentence frame such that the stress on the syllable was varied. As a result, the duration of the vowel varied; the shorter the vowel, the further from a target frequency the vowel formants fell. Lindblom was able to show that the failure of target attainment for the vowel could be predicted from a simple model. In the model, there is presumed to be a simple activating command for each of the three phoneme elements of the syllable. Due to such factors as the inertia of the articulatory structures, there is a time delay between the arrival of the command and movement completion. If successive commands arrive fast enough, the moving articulators will not attain target position. Lindblom also showed that the acoustic effects of increased speaking rate on vowel target attainment are the same as reduced stress.

Lindblom's model is based on the assumption, mentioned above, that stress does not affect the magnitude of vowel commands but merely their timing. Although he does not examine the question, it seems likely that he intends the undershoot mechanism to apply to consonants as well as vowels.

No electromyographic studies have been made, so far as I know, to test Lindblom's undershoot model directly. However, the prediction, by extrapolation, to the EMG level should be that, under conditions of varying stress, the EMG signals associated with phonemes remain constant but the spacing between them is altered. We do not have evidence about the behavior of vowels under varying stress, but we have performed a relevant experiment on consonants (Harris et al., 1968). When we compared the amplitude of electromyographic signals (orbicularis muscle) in words that were heavily emphasized in sentences with the same words unemphasized, we found that the stressed amplitude was some 10-20 percent greater. However, changes in lexical stress were not accompanied by amplitude changes. Fromkin in an earlier study (1966) has shown a larger amplitude in the orbicularis signals for initial /b/ closure in words than for terminal /b/ closure, although a related study (Harris et al., 1965) did not show a similar effect. From what we know about the way in which the EMG signal is related to the resulting articulatory motion, a larger amplitude
of EMG signal is translated into a faster acoustic transition and/or a more extreme articulatory position. Lindblom's model implies that this does not happen. It seems important to perform a direct EMG-analog of the Lindblom experiment and also to check the limits of the size effect. On the basis of present knowledge, a size change effect sometimes does occur. However, this may or may not represent an extreme maneuver, outside the range of the normal stress variation studied by Lindblom. Certainly, a system with signals of variable spacing but constant size is somewhat simpler to analyze than one in which signal size varies with stress and tempo of articulation.

Anticipatory Coarticulation

It has long been recognized that certain characteristics of a phone are likely to be anticipated in running articulation. For example, if a vowel is rounded, signs of the vowel rounding will occur in the articulatory gesture of the preceding consonant. Of course, this anticipation will act to produce different target configurations for the same phone in front of other different phones, with consequent acoustic and electromyographic effects.

This phenomenon appears to arise in a sequence of phones, the second of which is marked with respect to a characteristic that is unmarked in the first. For example, in English, the vowels are marked with respect to rounding, while many consonants are not. Therefore, such a CV sequence will show vowel rounding during the consonant. Similar types of anticipation during consonant production of aspects of the vowel are shown by MacNeillage and de-Clerk (1969). The determinants of the length of an anticipatory sequence are not known. Kozhevnikov and Chistovich (1965) have suggested that coarticulation boundaries act to delimit the syllable. In their model, the commands for a syllable are specified simultaneously with the start of the first command for which the commands are noncompleting, or at the syllable boundary. Consequently, there is variable coarticulation within the syllable but minimum coarticulation between syllables. Two recent experiments, by Daniloff and Moll (1968) and by Amerman et al. (1970), appear to contradict this view. Using measurements of X-ray film, they were able to show that coarticulation may extend over several phoneme units, even when word or syllable boundaries occur in the sequence. Some acoustic measurements made by Öhman (1966) of VCV sequences similarly show anticipatory coarticulation of the second vowel across the consonant, i.e., across the syllable boundary. He suggests that consonant gestures are somehow superimposed on slowly varying vowel targets. In all these examples, characteristics of the vowel are anticipated through a string of consonants. However, Amerman et al. cite an as yet (1970) incompletely reported experiment which shows anticipation of nasalization through a vowel sequence, again without regard to syllable boundaries.

No present model of articulatory behavior seems adequate to describe the circumstances and extent of anticipatory coarticulation. Syllable boundaries do not seem, in fact, to play the essential role that Kozhevnikov and Chistovich suggest. At least on the basis of present information, there does not seem to be the distinction of function between vowels and consonants in coarticulation that Öhman describes. Henke (1967) posits that each phoneme can be specified as a bundle of features, although each phoneme is not specified with regard to each feature. In the production of any sequence of phonemes, a high-level scan looks ahead from one positive value of a feature over a series of neutral values to the next specific positive value of the feature. However, Henke's description does not specify the time course
of anticipatory coarticulation in detail or, indeed, what characteristics of intervening phones affect it.

In short, then, although we are in a position now to specify some apparently general mechanisms that interfere with attainment of constant target, the scope of these mechanisms is not presently understood. One further question is how even approximate target maintenance is achieved. The usual explanation proposed (by MacNeilage, for example) is that the speaker keeps tract of articulatory position by feedback of some sort.

The Role of Feedback

Two types of feedback are clearly used at some stage in language development. The first of these is acoustic feedback. Acoustic feedback is clearly necessary for adequate speech development, as is demonstrated by the common failure of deaf children to speak intelligibly. Indeed, it has been suggested by Whetnall and Fry (1964) that children whose hearing is seriously impaired will speak normally if the remaining hearing is efficiently used. They suggest that the essential prerequisite to intelligible speech is an association between an articulation and some form of acoustic image, even if the acoustic image is seriously distorted.

Acoustic feedback has a somewhat more equivocal role in speech that has already developed. Noise masking of speech has little or no effect on speech intelligibility in adult speakers (see, for example, Ringel and Steer, 1963). Anecdotal experience suggests that traumatic deafness, in adulthood, does not cause immediate degradation of speech quality. On the other hand, continuous acoustic monitoring must have some role in speech maintenance; the speech of deafened adults does apparently deteriorate eventually, although this phenomenon has not been adequately studied. Furthermore, delayed auditory feedback has devastating effects on speech. In the delayed auditory feedback situation, the speaker is fitted with earphones and hears his own speech after a delay; the appropriate delay will cause slowing of speech, distortion of production, and stuttering (Fairbanks and Guttmann, 1958). In summary, then, continuous auditory monitoring appears to be unnecessary for speech production, but maintenance of normal articulation cannot survive serious distortion or deprivation of auditory feedback.

The second type of feedback believed to be important for the preservation of the integrity of speech is the feedback from the oral articulatory structures. This feedback may be one of two types, although they are not clearly distinguished in much of the literature. First, important feedback information may be conveyed by sensations arising from contact between oral structures, for example, the contact between the hard palate and the tongue tip. It is well known that the whole oral area is well endowed with sensory receptors for touch and pressure; Ringel (1970) has recently reviewed the literature on sensitivity of the oral region.

A second type of feedback that may be important in phoneme target maintenance is proprioception, defined by Scott (1970) as "the sensation which results in knowledge...of articulator position and movement which is traditionally believed to be mediated by muscle and jaw receptors," that is, information about the state of the muscle itself. Appropriate muscle receptors have recently been located in the laryngeal muscles by Baken (1969). The evidence for
appropriate receptors in the supralaryngeal muscles is summarized by Scott.

Assuming that such a mechanism is present, there is still the question of whether it is used in moment-to-moment monitoring of articulator position. Three types of evidence have been offered.

First, there is some clinical evidence that persons with reduced somesthetic perception will have deficient speech. MacNeilage et al. (1967) studied a patient with a grossly disturbed somesthetic system whose speech was almost wholly unintelligible. Cases of this sort, however, are difficult to classify. MacDonald and Aungst (1970) studied another patient who might be described as having the same pattern of deficit but whose speech was, to ear at least, perfectly normal.

Perhaps more convincing evidence comes from experiments on oral stereognosis. It has been shown (Shelton et al., 1967) that normal adults can recognize three-dimensional objects which are placed in their mouths. A recent study (Ringel et al., 1970) has shown that children with articulatory deficiencies are significantly less proficient at the oral stereognosis task. However, the oral stereognosis task is rather complicated, and its association with other types of performance has not yet been thoroughly studied.

A second type of evidence for the importance of sensory feedback in speech maintenance comes from examination of the effects of anesthetization of parts of the upper vocal tract. The classic study was performed by Ringel and Steer (1963) and has recently been repeated with some technical improvement by Scott (1970). In Scott's study, injection of topical anesthetic was believed to eliminate sensation from surface receptors in the tongue, palate, teeth, lips, and oral mucosa. Since the innervation of the kinesthetic receptors in the oral region is not completely understood, it is not known whether feedback from them was also blocked.

The chief effects of the multiple deprivation appear to be a reduction of the ability to refine labial and tongue blade and tip articulations, although speakers are able to maintain fairly intelligible speech. The interpretation of the findings as to feedback is, however, open to some question.

A recent experiment by Borden (reported by Harris, 1970) shows that, when the sensory block from the tongue is performed as in the experiments cited above, two of the muscles that may control positioning of the tongue tip can be paralyzed by spread of the injected anesthetic. It is not clear, then, the extent to which the results obtained are due to surface sensory deprivation, kinesthetic sensory deprivation, or motor paralysis.

The third type of evidence is from experiments where a significant correlation has been shown between successive aspects of the articulation of a syllable (MacNeilage, 1969; Ohala, 1970; Kozhevnikov and Chistovich, 1965). For example, the latter authors show a positive correlation between maximum jaw opening and the velocity of jaw opening and closing in the production of CVC monosyllables.

Kozhevnikov and Chistovich (1965) concluded, from the results of an earlier experiment: "Lowering of the...lower jaw must lead to a reflexive increase in
excitability of the centers of antagonistic muscles, while excitability must increase all the more with a greater drop of the...lower jaw." Ohala felt that "this...gives fairly good evidence of the presence and use of short-term feedback to make quick adjustments of articulator movement in speech" (p. 41). Again, the argument seems to be a little difficult. First, acoustic feedback, as well as kinesthetic feedback, was available to the subject, and second, one could argue that the results are accounted for by some general fluctuation in excitability affecting several successive gestures, rather than an actual feedback of one upon the next. One must agree with MacNeilage that the evidence for closed-loop control of speech is suggestive rather than definitive.

**PHYSIOLOGICAL REPRESENTATION OF THE PHONEME**

On the basis of our present knowledge, we appear to have failed to find a simple, absolutely invariant correlate of the phoneme at the peripheral levels thus far investigated.

The inertia of articulatory structures clearly limits the extent to which invariant positions can be attained on a phoneme-by-phoneme basis. Furthermore, there is considerable temporal uncertainty as to the time of application of parts of commands corresponding to a single perceptual phoneme entity. We should probably have anticipated the failure to find shape invariance, since attempts to find acoustic invariance, even for single speakers, have failed in the past. Clearly, if a listener extracts a string of invariant phonemes from the acoustic flow, he must do so on the basis of running perception of units that are larger than phoneme size, although strict syllable-by-syllable chunking does not seem to work very well either. Since there is no phoneme-by-phoneme progression in articulation, there is no need for precise closed-loop feedback control of articulatory position, nor is the evidence for such feedback very strong. Monitoring appears to be necessary only in some more general sense.

On the other hand, we do not appear to be very far away from appreciating the proper size of physiological segments. No one has proposed that articulatory context dependencies extend very far. If the syllable, as traditionally defined, is not the proper size unit, it is probably quite close to it. Only by careful, continuing study of articulatory coarticulation can we refine our knowledge of the field size over which perceptual scanning operates.

**REFERENCES**


Draper, M.H., Peter Ladefoged, and D. Whitteridge. 1958. Respiratory muscles in speech. JSHR. 2.16-27.


Flanagan, James L. and Lorinda L. Landgraf. 1968. IEE Transactions on Audio and Electroacoustics 16.57-64.


Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. JAcS. 27.765-768.


Harris, Katherine S., Malcolm M. Schvey, and Gloria Lysaught. 1962. Component gestures in the production of oral and nasal labial stops. JAcS. 34.743(A).


Houde, Robert A. 1967. A study of tongue body motion during selected speech sounds. Speech Communications Research Laboratory monogr. No. 2. Santa Barbara: Speech Communications Research Laboratory.


Sawashima, Masayuki, Thomas Gay, and Katherine S. Harris. In press. An electromyographic study of the pitch and intensity control mechanisms of the larynx. JSHR.


