Title
WPP, No. 18: A Phonetic Study of the Function of the Extrinsic Tongue Muscles

Author
Smith, Timothy S.

Publication Date
1971-08-01

Fromkin

A PHONETIC STUDY OF THE FUNCTION OF THE EXTRINSIC TONGUE MUSCLES

Timothy S. Smith

UCLA Working Papers in Phonetics
No. 18
August 1971

A phonetic study of the function of the extrinsic tongue muscles

Timothy S. Smith

Working Papers in Phonetics 18

May 1971

University of California, Los Angeles

ACKNOWLEDGEMENTS

This issue of Working Papers is a slightly revised version of a doctoral dissertation submitted to the UCLA Department of Linguistics in December 1970. This study presents some observations and data concerning the function of some of the extrinsic tongue muscles in normal speech, and attempts to relate the findings both to neurophysiologically-oriented models of the speech production process as well as to certain questions which have arisen in recent theory of generative phonology.

An experimental study of this sort cannot be accomplished at all without the assistance of many people. Dr. Thomas Calcaterra of the UCLA Health Sciences Center performed the anesthesia and electrode insertion in the second emg experiment. The X-ray film was made at Speech Synthesis Project laboratory, University of California, Santa Barbara. The willing cooperation of the late Prof. Pierre Delattre, and of Prof. André Malecot and Mr. Michael Melfi in matters cineradiological is gratefully acknowledged. Of all those who helped in the experimental work, I owe the greatest debt to Dr. Minoru Hirano, (Department of Laryngology, Kurume University, Japan). Dr. Hirano performed the technical manipulations for the first emg experiment with great skill, and was also very helpful in analyzing and interpreting the data.

I would also like to acknowledge the considerable work done by Kenneth Hill and Elite Olshtain (see UCLA Working Papers in Phonetics No. 1) in reviewing the literature concerning the anatomy and function of the tongue muscles. Their work was very helpful to me. Prof. Louis Goldberg was helpful in discussing some of the neurophysiological aspects of this work with me, and in providing additional references. To the technical staff of the UCLA Phonetics Laboratory, in particular Larry Grant, Richard Harshman, and Willie Martin, I owe many thanks for their assistance. Jeanne Yamane and Renee Wellin did a very skillful job of typing the final manuscript.

My work has profited from discussions with friends, colleagues, and teachers at UCLA, in particular George Allen, Kerstin Hadding, Frank Heny, Kay Atkinson King, Mona Lindau, John Ohala, Bill Riley, Marcel Tatham, Dale Terbeek, Ralph Vanderslice, and Harry Whitaker.

To Peter Ladefoged and Vicki Fromkin I offer my thanks for their excellent teaching, as well as considerable encouragement, help, and patience during my graduate studies, and during the (overly long) time that it took to prepare this dissertation. Finally, I would like to thank Barbara Alvarez for her help and continued encouragement.

Timothy Smith
Department of Linguistics
University of California, San Diego
La Jolla, California 92038

This research was supported by funds from USPHS (NIH) grant NB 04595, NSF grants GS 2741 and GS 1426, and the UCLA Department of Linguistics.

Chapter 1

Phonetics has traditionally been concerned with collecting, describing, and systematizing facts about the pronunciation of human language. Since pronunciation is most easily described in terms of positions and movements of the articulatory organs such as the lips, tongue, velum, and larynx, it is hardly surprising that most systems of phonetic transcription have been based upon configurations of the vocal tract. Nevertheless, there is still a considerable amount of confusion regarding the positions and dynamics of the articulatory organs of speech. For example, many of the investigations of the shape of the vocal tract in speech have been based upon lateral cineradiography (Moll 1960, DeClerk et al. 1965, Houde 1967, Perkell 1969). In implementing a line-analog speech synthesizer (Stevens et al. 1953, Rosen 1960, Umeda and Teranishi 1966, Hiki et al. 1968) it is necessary to know not only the cross-sectional diameter of the vocal tract, but also the area function; that is, the area of successive vocal tract cross sections. New techniques of experimental phonetics, such as high-speed cineradiology, ultrasonic scanning techniques (Kelsey et al. 1968), and electromyography are resolving some of the as yet unanswered questions. It is quite conceivable that about 10 to 20 years of reasonably intensive study of speech articulation will yield a very good over-all view of the articulatory aspects of speech production. We may know exactly what the lips, tongue, velum, larynx, and respiratory system do during connected speech, at least in certain types of normal careful speech.

But if we come to know all these things, will we necessarily know more about speech in general? Will we be able to construct better theories of phonetic representation, and will we know more about the mechanisms which speakers of a language use to control the actions of the vocal tract in connected speech? The answer, I feel, is quite obvious. No, not necessarily.

There has been considerable interest, in the past 20 or so years, in constructing and refining speech synthesizers. It is claimed that such devices are of considerable use and importance in speech research. This claim cannot be denied. The importance, for example, of speech synthesizers to research in the perception of speech is unquestionable. Most of the valuable work which has been done on determining the cues which enable hearers to identify certain segment-types (summarized in Liberman et al. 1967) would not have been possible without acoustic speech synthesizers. However, it has been claimed that synthesizers are much more than mere tools, useful in the study of speech perception. Kim (1966), Ladefoged (1967b), Mattingly (1967), Tatham (1970), and others have claimed that a speech synthesizer is useful as a model of speech production. That is, given a fixed input string representing some abstract linguistic units (phonemes or allophones), and an algorithm for converting these input units into an acceptable acoustic signal, have we then achieved a successful model of the speech production process? The answer--again--is no, not necessarily. The synthesizer is a true model of the speech production process only to the extent that it incorporates the constraints involved in the speech production process as it is carried out by humans. We may have achieved some insights into part of this process but, unless we understand something of the neurophysiological constraints involved in the speech process, the model will be incomplete.

In order to make these statements more explicit, we will mention some of the problems involved in constructing various types of synthesizers. Consider the following types of vocal-tract synthesizers (we disregard here the synthesis of the laryngeal source waveform).

Synthesizer A: This device takes successive area functions of the vocal tract over time (at, say 5 msec intervals) and converts them into a time-varying speech waveform. A successful synthesizer of this sort is a model of the transfer function of the vocal tract. Its implications are for the testing of the acoustic theory of speech production (Fant 1960, Stevens and House 1961, Flanagan 1965). The model has implications for the total speech production model if the area functions are non-arbitrary, that is representative of the actual area functions in speech.

Synthesizer B: This synthesizer does everything that A does, but in addition derives the area function from successive mid-sagittal outlines of the vocal tract. Work on synthesizers of this sort has been carried out (Flanagan et al. 1970, Rice 1971).

Synthesizer C: In this synthesizer, the mid-saggital outlines of the vocal tract are derived from some more abstract representation or parameterization. Synthesizer B is conjoined to produce the acoustic waveform. It is at the point of constructing synthesizer C that most models of the speech production process run into trouble. It is feasible (although by no means easy) to construct a model which will store idealized vocal tract shapes corresponding to some discrete linguistic representations, and which will generate successions of mid-saggital vocal tract shapes by applying certain functions to the stored configurations. Henke (1966) has constructed a model of this sort. However, we do not yet know whether this is what a speaker of a language does when he produces intelligible speech. It could equally well be proposed that the speaker has stored in his permanent memory certain patterns of muscular activity, and that when he speaks he constructs an abstract phonetic representation by concatenating these stored patterns and issuing the commands to the muscles of speech. (This topic will not be pursued in detail here, since we will return to it in Chapter 4.)

In either case, the speaker must move his muscles in order to speak. Even if we assume that the underlying units of a speech production model are idealized representations of vocal tract shape, the speaker must transform these representations into muscular activity in order to produce sounds. This transformation process would be uninteresting if it were totally idiosyncratic. This, however, is very unlikely. Human anatomy differs little from individual to individual. All normal individuals have a certain fixed arrangement of the muscles in the facial region, tongue, velum, larynx, and thoracic cavity. These muscles are innervated by the same nerves, which originate from the same place in the brainstem or spinal cord, for each individual. It is improbable that the processes by which the muscle activity is generated are markedly different from one individual to the next.

It is clear that it will not be possible to construct a synthesizer of type C that is an accurate model of the speech production process given our current lack of knowledge concerning this process. In other words, we need to do a considerable amount of additional analysis before we can hope to do any physiologically or linguistically meaningful synthesis. The reservations concerning the meaningfulness of synthesizers of Type C apply, even more strongly, to direct acoustic synthesis (terminal-analog synthesis). Terminal-analog synthesizers have, in the past, produced the highest quality speech, and are, as we have mentioned, quite useful in perceptual studies of speech. However they are in no sense accurate models of the human speech production process.

These comments should not be construed as depreciating the work (much of it excellent and quite impressive) which has been done on line-analog and terminal-analog synthesis. Nevertheless, it is necessary to keep it in perspective when discussing synthesis of speech as a model of speech production.

The relation between theories of language structure and theories of speech performance is a much discussed subject. Much of this discussion, however, is polemical, and of questionable value. The meaningful issues are more often obscured than clarified. Despite the pitfalls, some discussion of this topic might prove to be useful in clarifying the intentions of this study.

Basically, a theory of language structure attempts to state generalizations about language *per se* by formally relating a less abstract level of structure to a more abstract level. The constraints on the relations are derived internally, and are not directly related to external criteria. The generalizations handle facts which are basically intuitive in nature (for example, judgements of the grammaticality of sentences). Ultimately, a complete theory of language must state the possible relations between meanings and sound; that is, between the idea of a speaker and his articulation of the idea. Some of these relations will be universal, and some will be specific to the speaker's language.

Within a theory of language, the statement that a unit X at some level is related by a process Y to a unit Z at some less abstract level does not imply that the behavioral correlates of X, Y or Z are stated in the theory. There is one possible exception to this observation. That is when the unit Z happens to be a unit at the phonetic level. The status of units at this particular level of representation is highly problematical.

In the theory of language usually called "transformational grammar," the least abstract level of linguistic representation is the level of systematic phonetics (Chomsky 1964, Postal 1968, Chomsky and Halle 1968). The treatment of this level of structure is curiously paradoxical, however. On the one hand, it is claimed that representations at this level are mentalistic. They are not to be considered as a faithful, detailed transcription of the speech signal, but rather as a statement of the speaker's "knowledge" of the pronunciation of the utterance (Chomsky and Halle 1968, p. 294). On the other hand, the features in terms of which this level is represented appear to be very "phonetic" in the narrow, transcriptive sense. Most of the features described by Chomsky and Halle are binary articulatory features which describe the state of the vocal tract in speech. As Ladefoged (1970, p. 28) has pointed out "nearly all their [Chomsky and Halle's] arguments for or against a particular system of universal features are based on observation of what goes on in particular languages. Nothing crucial is based on evidence concerning child's language learning ability, nor on our current neurophysiological knowledge concerning his innate endowment."

As implied in Ladefoged's statement, there are a number of criteria which may be relevant to the selection of a set of phonological/phonetic features, and which deserve systematic study. In this study, we will be concerned with the implications of physiological phonetics for the evaluation of a phonological/phonetic feature system. These questions will be raised and discussed in greater detail in Chapter 4.

A linguistic theory of phonological processes will map abstract underlying representations (e.g., systematic phonemes in generative phonological theory) into another level of structure (systematic phonetics) which is abstract, but yet closer to the pronunciation of the utterance involved. A speech production model also works in the same general way, that is it maps one level of structure into another. Consider synthesizer A discussed above. This device models the transfer function of the vocal tract, and converts articulatory shapes into an acoustic waveform. It is clearly a sub-part of a total speech production model. The major difference between a speech production model and, say, the phonological component of a grammar is that if a speech production model maps a representation A into a representation C by a process B, then at least the representation C and the process B must be physically real. In synthesizer A, both levels of representation are physically real and are describable in physical terms. Furthermore, the process which relates them is given by a study of the physics of sound in certain air-filled, highly non-uniform resonators.

In synthesizer C above, which derives vocal tract shapes from an abstract representation, the output (the vocal tract shapes) is clearly physically real and is, ideally, observable and describable. The more abstract representations may also, ideally, have some neurophysiological or psychological reality. But it is crucial that the "transfer function" also be real in some sense. We could require that it have a strong degree of reality; that is, that it have neurophysiological correlates. Or we could require only a "weaker" degree of reality, such as that it be psychologically real.

Assume that a model of synthesizer C is constructed which does indeed convert abstract units into vocal tract shapes by a process which is correlatable with (or is isomorphic to) a neurophysiologically real process. Does this model represent a performance model relatable to a competence model such as that provided by a generative phonological description? This is a difficult question to answer directly, but I think that the answer is probably that it does not. The partial model of linguistic competence provided by a generative phonological description is not a model which has as its main function the generation of actual utterances. That is, this model does not represent a conversion of abstract signals into actual articulatory gestures. What it does do, rather, is to state linguistic generalizations concerning sound structure by relating two levels of possible sound structure. There is little reason to suppose that the processes involved in this relation will have any relation to the processes involved in the actual production of speech.

In general, what we are saying is that a complete model of speech production would not necessarily have the structure implied by the diagram shown in Figure 1.1. A more likely model is that shown in Figure 1.2 (cf. also Fromkin 1968a, Laver 1968). This model is not intended to imply that phonology is subservient to, or a peripheral part of, speech production. It does claim that the generalizations stated in a phonological theory may not be a central part of the speech production process.

In Chapter 4 we will postulate a partial model of the process which converts abstract phonetic representations into speech. We will be concerned primarily with the conversion of the abstract features into muscular movements. In discussing this model, we will assume that it has as its input the abstract phonetic representations, which are the output of the phonological rules. Whether or not these rules are actually "run through" in the production of the utterance is highly debatable. Fromkin (1970) notes that evidence from speech errors indicates that mistakes in serial ordering follow the phonological pattern of the language. For example, she observed that in the utterance "tab stops," when the voicing of the labial stops was metathesized, the final segment was also altered, so that the phonetic output was [tʰæpstɑ·bz] rather than [tʰæ·pstabs]. The processes which lengthen vowels and which given the proper phonetic shape to the plural suffix had operated within the metathesized phrase. This evidence indiates that processes similar to phonological rules are involved in the actual generation of sentences.

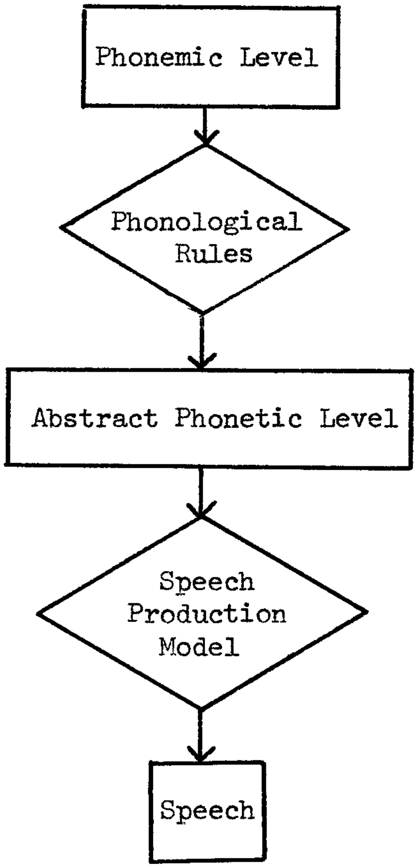


Figure 1.1

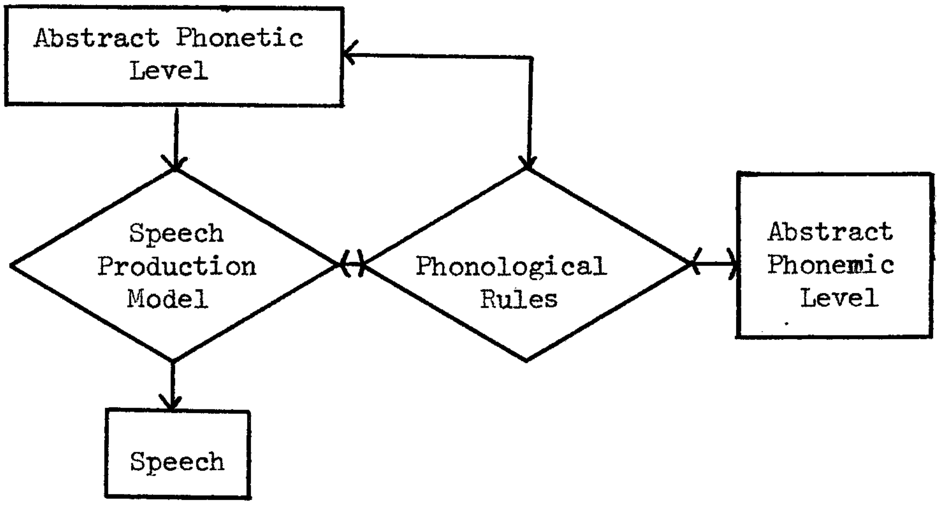


Figure 1.2

Clearly, there is much basic data which must be collected before viable articulatory models of the speech production process can be postulated. We are just beginning to understand the dynamics of muscle activity in speech. Activity of the facial muscles has been explored (Harris et al. 1965, Fromkin 1966, Öhman 1967b, Tatham and Morton 1969, Harris et al. 1969), and the activity of the intrinsic muscles of the larynx is now beginning to be understood (Hirano et al. 1967, Ohala and Hirano 1967, Ohala 1970). Velar function has been very productively studied (Lubker and Moll 1965, Fritzell 1969). Of all the "organs of speech" however, the tongue is clearly the most important. Nevertheless, systematic studies of tongue muscle function are very rare. Ladefoged (1964), MacNeilage and Sholes (1964), and MacNeilage and DeClerk (1969) are the only extensive studies of this aspect of speech production from the point of view of specific muscle function.

The purposes of this study are twofold: (1) to gain some basic information concerning the function of individual extrinsic tongue muscles in normal speech, and (2) to explore the implications of this data for models of speech production and theories of language structure. The first of these is quite straightforward. The results of our investigations are reported in Chapter 2. The second aspect of this study, however, is necessarily quite preliminary. In Chapter 3, the emg data from the extrinsic tongue muscles are considered from the point of view of speech production models. Most investigators who have collected emg data on speech events have noticed that the patterns of muscle activity differ considerably for instances of segments which are identical at some more abstract level. That is, the concept of invariance of motor commands with respect to abstract units has been consistently disproved. We noted this effect also. The crucial question, however is whether this observed variability can be predicted by considering the ways in which peripheral neurophysiological feedback mechanisms can modify the motor commands. We are essentially in agreement with MacNeilage and DeClerk (1969), who have argued that not all instances of observed variability of emg signals can be accounted for by peripheral feedback mechanisms.

In the fourth chapter, we discuss the implications of the physiological phonetic research reported in Chapters 2 or 3 for theories of language structure. We consider some processes that might be involved in converting abstract phonetic representations into motor commands to some of the extrinsic muscles of the tongue for a limited set of phonetic segments. The implications of the model proposed for units and processes of theories of the language structure is discussed.

Chapter 2

In this chapter, the methods and results of two electromyographic experiments on the muscles of the tongue will be presented. In discussing the results of these experiments, we will focus on the role played by the individual muscles in shaping the tongue during speech production. Aspects of the results which are significant from the point of view of the control of speech production will be discussed in the later chapters.

ANATOMY OF THE TONGUE

Much of the discussion in this and the following chapters will require specific reference to details of the anatomy of the tongue. For this reason a brief description of tongue anatomy, particularly with regards to the musculature, is being included. The muscles of the suprahyoid group are also described. This description is intended as an aid to the reader, not as a scholarly or definitive description of the anatomy of the tongue and surrounding areas. In preparing this account, I have drawn from a number of sources, principally Spalteholz (1933), Abd-el-Malek (1939), Dabelow (1951), Strong (1956), Van Riper and Irwin (1958), Pernkopf (1963), Ladefoged (1964), and Fritzell (1969).

The tongue is composed mostly of muscular tissue. This, of course, accounts for its considerable flexibility and mobility. In anatomical descriptions, it is usual to refer to "upper" surface of the tongue as the dorsal surface. The direction opposite to dorsal is ventral. Anterior, of course, means towards the tip, and posterior towards the root. The area around the mid-line of the tongue body will be called the medial portion, the edges of the tongue the lateral portions. The most prominent feature on the dorsal surface is a V-shaped line of papillae (large buds), which are mushroom shaped if closely examined. These lie about two thirds of the way from the tip to the root of the tongue. The line is called the terminal sulcus. These features can be seen in Figure 2.1, a drawing of the dorsal surface of the tongue.

Phoneticians have traditionally divided the tongue into the tip, the blade, and the dorsum. The most posterior portion of the dorsum is called the root. In the descriptions in this chapter, it will be more convenient to talk of the tongue as if it were divided into three approximately equal sections: the anterior, medial, and posterior portions (or thirds of the dorsal surface). The posterior portion can be defined as the area below the terminal sulcus and above the root of the epiglottis. The anterior and medial thirds are separated by an imaginary

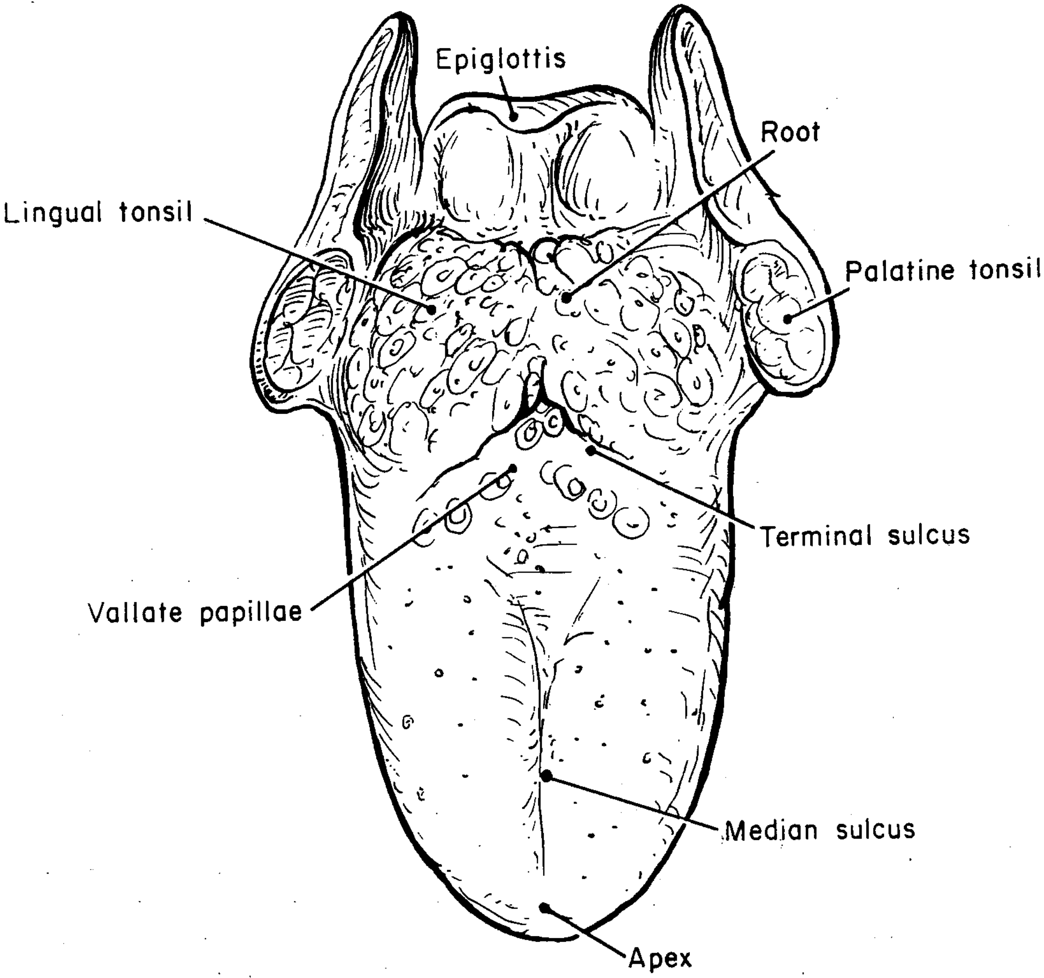


Figure 2.1

Dorsal surface of the tongue. After Spalteholz (1933).

line halfway between the terminal sulcus and the tip. These divisions are, of course, approximate and somewhat arbitrary. Nonetheless, they provide convenient reference points for discussion of the muscles of the tongue.

These muscles are normally divided into two groups: the extrinsic muscles and the intrinsic muscles. Many muscles in the human body can be described in terms of a point of origin, and a point where the muscle fibers insert into other fibers, or onto a hard structure such as a bone or cartilage. The tongue muscles which originate outside the body of the tongue on skeletal structures are the extrinsic muscles. These muscles insert into the body of the tongue, and in general can be thought of as controlling the position of the tongue body in the oral cavity. As we will see below, they can also be used to change the shape of the tongue to a great extent in speech. These extrinsic muscles insert into the intrinsic muscles of the tongue, into other extrinsic muscles, onto septa within the tongue, and into the membranous surfaces of the tongue. The intrinsic muscles of the tongue cannot be conveniently described in terms of a point of origin and a point of insertion, as they are contained wholly within the body of the tongue. These muscles are named with reference to the principal axis of their fibers: longitudinal, vertical, or transverse.

The tongue contains no hard structures, i.e. cartilages or bones. Just below the tongue body itself is the hyoid bone, a structure roughly the shape of a horseshoe, with two sets or protrusions, called *cornua* or "horns". The hyoid bone is unique in that it is the only bone in the human body which does not connect directly to another bone. The hyoid bone forms the base for the tongue, and is a point of origin or insertion for several important muscles which can control the position and shape of the tongue. The hyoid bone attaches to the epiglottis by means of the hyoepiglottic ligament, to the thyroid cartilage by means of the hyothyroid ligament, and to the styloid process of the temporal bone by means of the stylohyoid ligament. Its muscular attachments will be described below. The hyoid bone and some of its attachments can be seen in Figure 2.4.

The anatomy and probable function of the extrinsic and intrinsic muscles of the tongue is described below. It should be emphasized that the standard descriptions of the function of individual tongue muscles are based almost entirely upon anatomical considerations. There is very little solid physiological evidence concerning the functioning of individual muscles in the human tongue. Evidence of this sort is somewhat difficult to gather, since individual tongue muscles do not effect movements of hard structures which would be observable by X-ray techniques, as is the case in the larynx and in most other structures used in speech. Perhaps the only really feasible procedure for collecting data on the actions of individual tongue muscles is electromyography. In any case, anatomical considerations can provide some very good hypotheses concerning the probable actions of these muscles, hypotheses which can be evaluated by electromyographic and other procedures.

Intrinsic muscles of the tongue

Superior longitudinal muscle: this muscle lies immediately below the dorsal surface of the tongue. It runs from the tip of the tongue back to a point about two thirds of the way from the tip (near the terminal sulcus). Most of the intrinsic and extrinsic muscles of the tongue are paired muscles, that is, there are two portions -- right and left -- to the muscle. Some anatomists (Spalteholz 1933) consider the superior longitudinal to be the only unpaired muscle of the tongue. It is thickest in the medial portion of the tongue. Its probable function is to retract the tip of the tongue (MacNeilage and Sholes 1964) and to elevate the tip in the production of dental, alveolar, and alveolo-palatal consonants, as well as retroflexed vowels.

Inferior longitudinal muscle: this is a paired muscle which lies on either side of the median septum. Fibers of this muscle run from the most anterior portion of the tongue back to the mucuous membrane of the root of the tongue. This muscle can act as an antagonist to the superior longitudinal, depressing the tip of the tongue after it has been elevated. It can also act synergistically with the superior longitundinal to retract the tip.

Transverse muscle: consists of fibers which fan out in a roughly horizontal direction from the median septum of the tongue towards the lateral surface. Dabelow (1951) claims that fibers from the transverse and verticle muscles do not originate at the median septum of the tongue, but that they actually may cross the middle of the tongue and insert into other intrinsic or extrinsic fibers. He also claims that in some places the median septum is not evident at all, but is no more than a place where fibers of the transverse muscles meet and sometimes cross. This muscle is important in narrowing the tongue, as when the tongue is protruded with a pointed tip. Strong (1956) claims that this muscle is active in deforming the anterior portion of the tongue to produce the precise alveolar or dental contact found in coronal consonants.

Verticle muscle: fibers from this muscle run in a perpendicular direction to the fibers of the transverse muscles. They run ventrally and laterally from the dorsal surface of the tongue to the sides and ventral surface. The function of this muscle is to flatten the tongue, as in protrusion with a "spread" tongue. Its function in speech is not clear, although it may be active in sounds produced with an open vocal tract, such as low vowels.

In general the intrinsic muscles are found in the anterior and medial portions of the tongue, although a few fibers from the longitudinal muscles may interdigitate with extrinsic fibers in the posterior one third of the tongue. Fibers from the intrinsic muscles insert onto the dorsal and ventral membranes of the tongue surface. This may account for the fact that the membrane of the medial and anterior portions cannot move, or slip, independently of the muscle immediately below it, while in the posterior part of the tongue the membrane can so move. The intrinsic muscles are difficult to visualize, and very difficult to see consistently during dissection. This is mainly because the fibers from these muscles become interlaced with fibers from other intrinsic muscles, and from extrinsic muscles inserting into the tongue. These muscles are easier to see in the tongue of a fetus or a new-born infant. Figure 2.2 shows the intrinsic muscles of the tongue.

Extrinsic muscles of the tongue

Genioglossus: this large muscle originates at the symphasis menti on the interior surface of the front part of the mandible, and inserts into the tongue in a fan-shaped manner (see Figures 2.3 and 2.4). Fibers of this muscle run towards the most anterior part of the tongue, towards the medial portion, and towards the root. In later discussion, I will refer to the anterior and posterior portions of this muscle, since electromyographic evidence indicates that this muscle should properly be considered to be composed of several separate and independent muscles, which may be independently controlled. In general, the genioglossus serves to pull the tongue forward in the oral cavity. Contraction of the posterior fibers alone can serve to widen the pharynx and "bunch" the tongue, as in the production of a high vowel like [i] or [u], or a velar stop or fricative.

Hyoglossus: the hyoglossus is a thin, quadrilateral muscle which originates at the upper border of the greater hyoid cornua and from adjacent portions of the hyoid body. The fibers run upwards and slightly forward and insert into the lateral margins of the tongue. The hyoglossus fibers decussate with the horizontal fibers of the upper and lower parts of the styloglossus muscle. Some fibers originate at the lesser hyoid cornua, and these are termed by some anatomists the chondroglossus muscle (Spalteholz 1933). The action of the hyoglossus is to depress the body of the tongue. It is probably very active in sounds involving an open vocal tract, such as the low front and back vowels [æ], [ɑ], and [ɔ]. It is also probably very active in the production of pharyngeal ([ʕ]) and pharyngealized consonants.

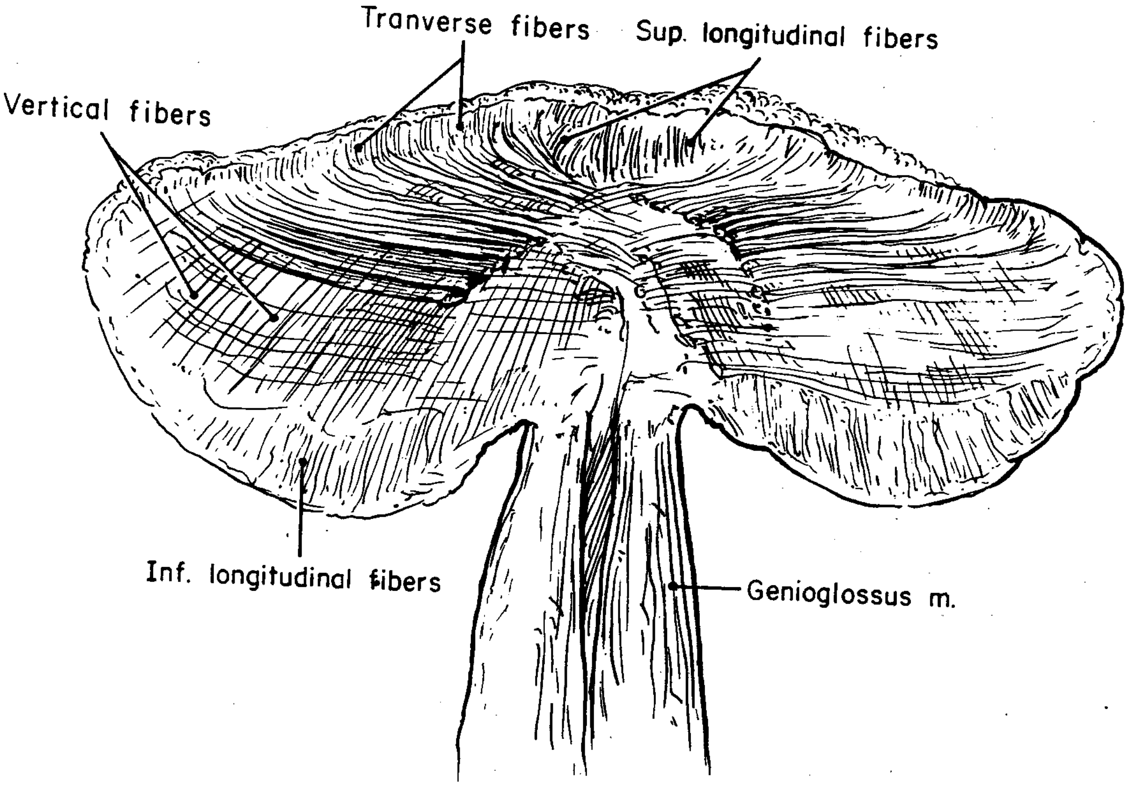


Figure 2.2

Frontal section in the anterior third of the tongue, showing the intrinsic muscles. Fibers of the genioglossus muscle are also shown. After Strong (1956).

Styloglossus: the styloglossus,as its name implies, originates at the styloid process of the temporal bone, and from the anterior part of the stylohyoid ligament. Its fibers run downward along the side of the upper part of the pharynx, then curve in an anterior direction, inserting into the sides of the tongue. Some fibers from the styloglossus also form a small bundle which decussate with the fibers of the hyoglossus mucle (Figure 2.4). The styloglossus appears to be the major muscle involved in retracting the medial and posterior parts of the tongue. It would appear to be active mostly in the production of high back vowels and velar consonants.

Palatoglossus: this small muscle, about 1.5 by 3.0 mm at its midpoint (Fritzell 1969), originates at the palatal aponeurosis, and inserts into the sides of the tongue, somewhat anterior to the major point of insertion of the styloglossus fibers. It forms the anterior faucal pillars of the oral cavity. It is sometimes considered to be a muscle of the palate, in which case it is often named the glossopalatine muscle. This muscle can act as an antagonist to the levator veli palatini muscle, serving to lower the velum. It is also active in certain speech sounds, although its small size relative to the mass of the tongue would indicate that it probably contributes little to positioning the tongue. This muscle is sometimes damaged during tonsillectomy, and the patients almost invariably regain normal speech.

Glossopharyngal muscle: this muscle is a part of the superior pharyngeal constrictor. The superior pharyngeal constrictor forms a loop open in the anterior dimension. The superior fibers of this muscle insert onto the pterygoid plate, the hamulus, the pterygo-mandibular raphe, and the interior of the mandible posteriorly. The inferior fibers are the glossopharyngeal part, and they insert into the sides of the tongue near the terminal sulcus. The glossopharyngal muscle would probably be most active in producing sounds with pharyngeal constriction, such as [ʕ], and low vowels such as [ɑ] and [ɔ].

Suprahyoid muscles

These muscles control the placement of the hyoid bone. They are used to elevate the hyoid, retract it, and draw it upwards and forwards, and to stabilize it against contractions of the infrahyoid muscles such as the sternohyoid and the omohyoid.

Mylohyoid: originates at the mylohyoidal ridge on the interior part of the mandible. The anterior fibers of this muscle run from one side of the mandible to the other, the posterior fibers insert onto the anterior portion and sides of the body of the hyoid bone (Figure 2.2). This muscle forms a sling under the tongue body, and can be used to lift the tongue in certain circumstances. It should be active in high vowels and in lingual consonants.

Geniohyoid: originates at the symphasis menti of the interior front surface mandible, and inserts onto the anterior part of the hyoid bone. May be used to pull the hyoid bone forward and slightly upward, and to assist in opening the jaw when the hyoid is fixed by antagonistic muscles. The geniohyoid muscle is shown in Figures 2.3 and 2.4.

Digastric: consists, as its name implies, of two bellies, the anterior and the posterior. These are joined by a piece of tendon which attaches to the hyoid bone. The main function of this muscle, the most inferior in the group, is to open the jaw. It is claimed to be more active at extreme degrees of opening (MacConaill and Basmajian 1969).

Stylohyoid: originates at the styloid process of the temporal bone, and runs obliquely downward, inserting onto the greater cornua of the hyoid bone. Can be used to elevate and retract the hyoid bone (Figure 2.4).

All of the muscles of the tongue itself, with the exception of the palatoglossus, receive their motor innervation from the XIIth cranial nerve (the hypoglossal). The sensory innervation of these muscles will be discussed in detail in Chapter 3.

There are several standard methods which can be used to obtain information concerning tongue movement in speech. Palatography, both static and dynamic, provides information concerning the contact points between the dorsal surface of the tongue and the teeth, alveolar ridge, and hard palate. Such information is useful in determining degree and shape of lingual contact in the production of certain obstruents, but gives no information on shapes of the tongue during vowels (there is some dental and palatal contact in very high front vowels, though). Most of our information concerning vocal tract shapes for vowels has been derived from X-ray studies, direct observations and casts of the oral cavity, and simple kinesthesis. These techniques, used singly or in combination, are somewhat limited,however.

There has been very little systematic electromyographic (emg) investigation of the tongue muscles. Reasons for this state of affairs are not difficult to find. From a purely medical-pathological point of view, the tongue is of little interest. Diseases of the tongue are relatively rare, and in extreme form are usually treated by total glossectomy.*

---

* This invariably results in a loss of normal speech, however it has been reported (Goldstein 1940; Drachman 1968) that intelligible speech can be produced by glossectomized persons, after a period of recovery and relearning. This is not particularly surprising, when the redundancy of the speech code is considered. We are normally able to "understand," if not recognize in the narrow sense, speech produced by some very strange mechanisms:

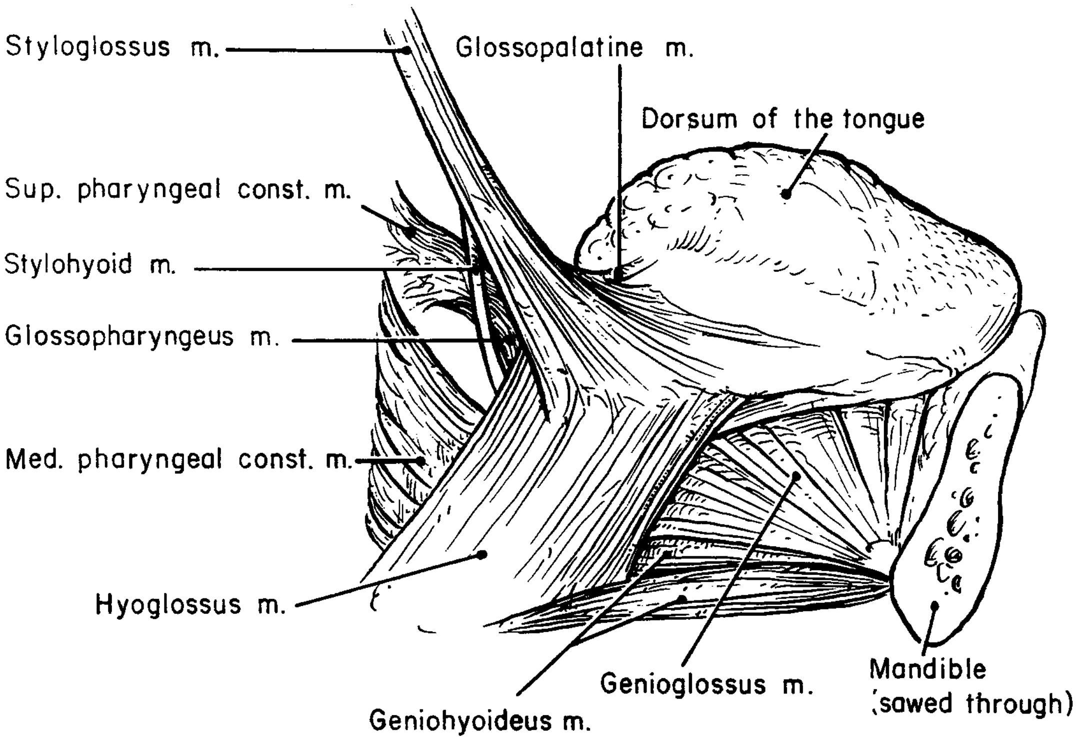


Figure 2.3

Extrinsic muscles of the tongue and suprahyoid muscles. The glossophrayngeal muscle is shown behind the stylohyoid. After Spalteholz (1933).

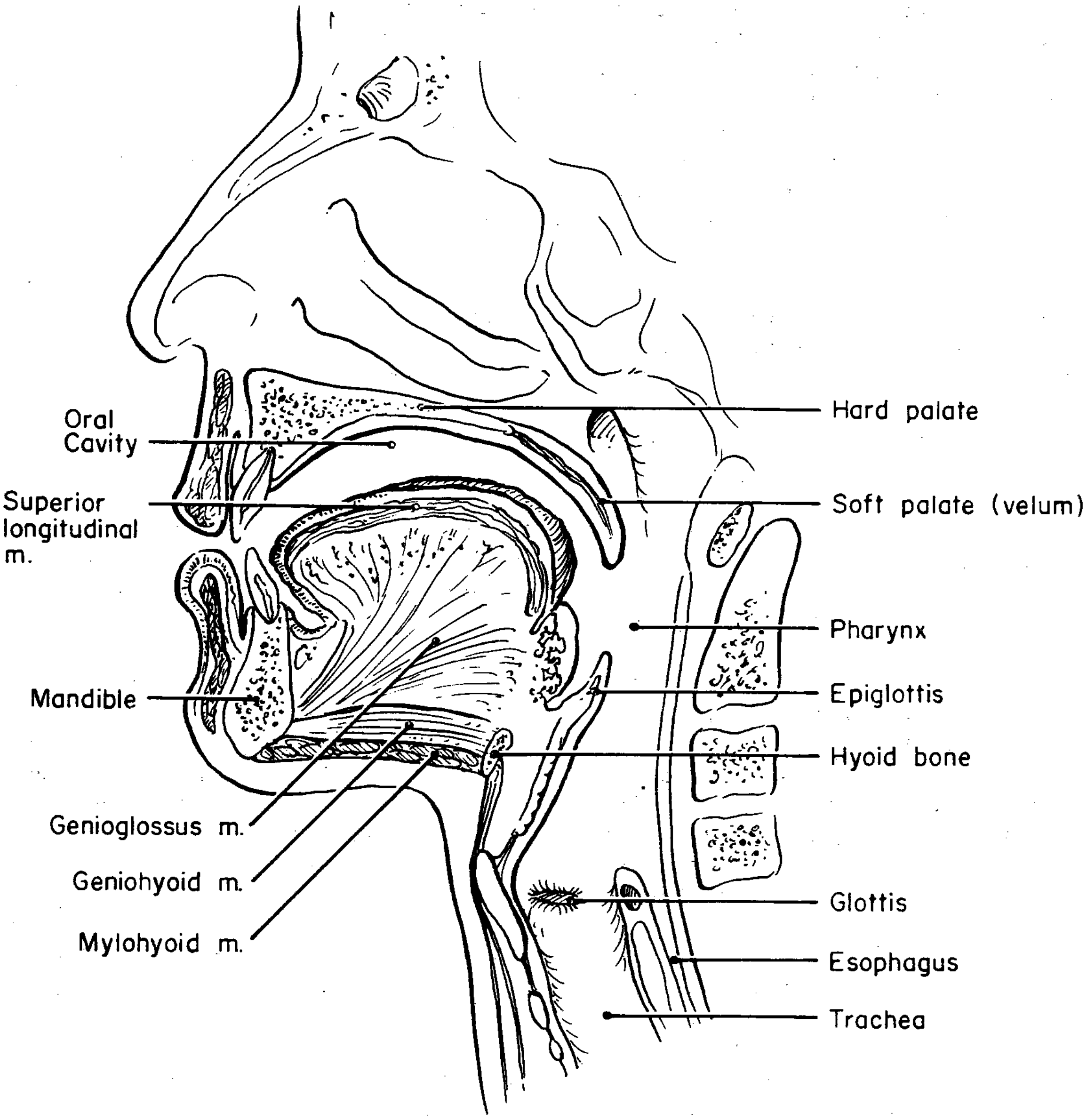


Figure 2.4

Midsaggital section through the vocal tract, showing some of the intrinsic and extrinsic muscles of the tongue, and the infrahyoid muscles. After Pernkopf (1963).

Perhaps of more importance in explaining the apparent disinterest in the specific muscular functioning of the tongue is the fact that these muscles are very difficult to isolate. It is all but impossible to record unambiguous emg potentials from the intrinsic muscles of the tongue, since these muscles do not consist of large fascicles of fibers. The fibers of the intrinsic muscles interdigitate with fibers of other intrinsic muscles, and with extrinsic muscle fibers. Some investigators (Cosi and Tonali 1966) claim to have recorded from intrinsic muscles, however their claims should be regarded with some skepticism. The extrinsic muscles of the tongue are more accessible, but even here rather special techniques are called for. Surface electrodes will seldom record from single extrinsic muscles unambiguously, since many of these muscles lie well below the skin or mucuous membrane. Ordinary needle electrodes are uncomfortable, and inhibit many of the movements called for in normal speech. Special fine wire electrodes can be used, however, to record emg signals from these muscles.

A number of authors have studied emg signals from the tongue (Mashiko 1960; Cosi and Tonali 1966; Caraceni and Zibordi 1966; Cunningham and Basmajian 1968); however they did not investigate the activity of specific isolated muscles in speech, and their findings will not be reviewed here.

Bole and Lessler (1966) investigated the function of the genioglossus muscle in man using bipolar wire electrodes inserted into the muscle near the frenulum linguae. They found that the genioglossus was about twenty times more active isometrically, as in attempting to protrude the tongue against the resistance of the upper teeth, than in isotonic contraction, as in protruding the tongue between the lips. Bole and Lessler did not study the action of this muscle in speech in any systematic fashion. Their subjects pronounced the numbers one, two, three, ... ten. They found that activity was greatest in the pronunciation of the words "three", "eight", and "nine". Since they did not publish a simultaneous audio signal, it is not possible to decide with which segments in these words the muscle activity should be associated. Judging from our investigations, as reported below, it seems likely that the activity should be associated with the high front vowel [i] which is present in each of these words, either directly or as a component of a diphthong. This allows us to infer that Bole and Lessler were recording primarily from the posterior fibers of the genioglossus. This in turn probably means that the activity of attempting to protrude the tongue against an obstruction is not true isometric muscle activity, since under these circumstances the posterior fibers of the genioglossus will certainly shorten and increase the width

---

the synrinx and pharynx of the India Hill mynah; "Donald Duck" speech, produced by lateral tongue frication modulated by the buccal cavity, teeth, and lips; or the distortions in normal speech produced by speaking with a mouth full of food. The fact that some patients who have undergone total glossectomy can recover to the extent of producing intelligible speech does not, as Goldstein implies, indicate that the tongue is unimportant in speech production.

of the pharynx.

Faaborg-Anderson and Vennard (1964) recorded emg signals from several extrinsic laryngeal and supra-hyoid muscles during singing in different registers. One of their results was the finding that the mylohyoid was active in the phonation of the high vowels [i] and [u], but not in the low vowel [a]. One subject, a native speaker of Danish, showed marked mylohyoid activity in producing the vowels [i, y, u], and little or no activity for this muscle in producing [ɑ, o, ø]. Their results were for steady-state, sung vowels, however, and thus are not typical of normal speech activity.

The most extensive electromyographic investigations of the tongue in speech are those of MacNeilage and Sholes (1964) and MacNeilage and DeClerk (1969). In the 1964 study, emg signals were recorded from 13 locations on the dorsal surface of the tongue, between the median line and the lateral edge. Surface electrodes were used, and they were placed from the tip (location 1) back to the base of the root of the tongue (location 13). Activity during the production of 17 vocalic nuclei in the environment /p _ p/ was examined. Their results can be summarized as follows:

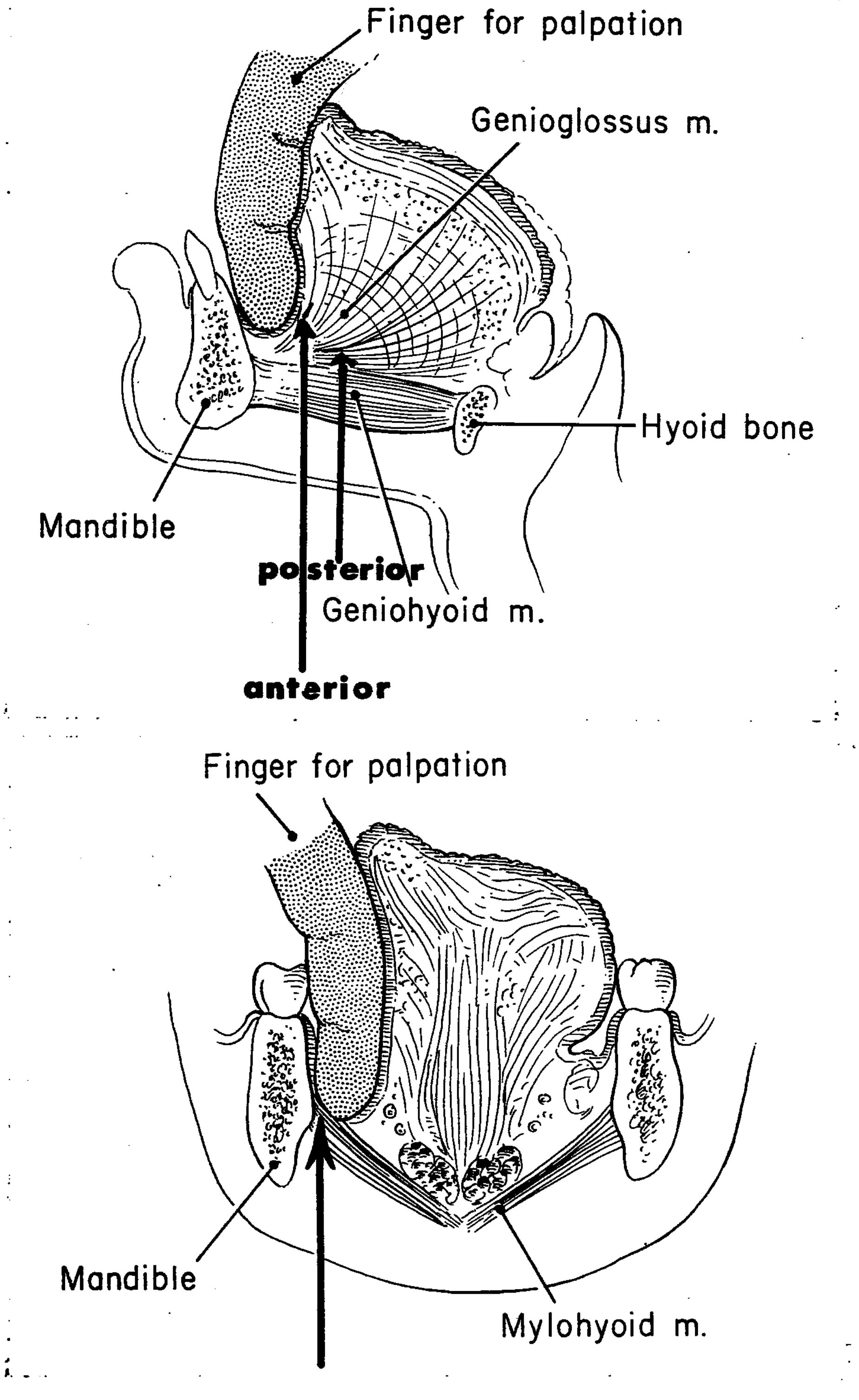
1.) Activity from the 3 most posterior electrodes was noted in the production of nuclei involving a high front tongue position ([i, ei, ɑi, oi]). Such activity was not observed in the vowel [ɪ], nor in high back vowels such as [u].

In the case of the vowel [i], the peak was located at the onset of voicing. For the other vocalic nuclei, the peaks were located during the voicing, later than for [i]. The authors attribute these peaks to contraction of the posterior fibers of the genioglossus muscle, pulling the tongue forward for [i] and the glide component of [ei, ai, oi].

2.) A large amount of activity for the most posterior 4 electrode locations was found coincident with the onset of voicing for the vowels [ʌ, a, ɔ]. The authors considered this activity to be a result of contraction of the hyoglossus in conjunction with the styloglossus.

3.) They claimed that the superior longitudinal muscle may be acting in conjunction with the styloglossus at any time the styloglossus contracts.

4.) Amount of electrical activity was, in general, proportional to tongue movement. Greater deviation from a hypothetical rest or neutral position in the vocal tract produced correspondingly higher voltages at the recording electrode.



Figures 2.5 (top) and 2.6 (bottom)

Figure 2.5 shows the method in which electrodes were inserted into the genioglossus muscle. Electrodes were inserted into two locations in this muscle, as shown in the illustration. The bottom figure (Figure 2.6) shows the insertion into the mylohyoid muscle. In this muscle, the insertion was made anterior to the hyoid bone.

The study by MacNeilage and DeClerk (1969) also involved the recording of emg action potentials from the tongue. Since they were mainly concerned with coarticulation effects between consonants and vowels, detailed consideration of this article will be deferred until Chapter 3.

EMG EXPERIMENT #1

In order to determine something of the function of the tongue muscles in speech, we decided to sample emg activity from some of the more accessible extrinsic muscles of the tongue and suprahyoid area. The muscles selected in this case were the genioglossus and the mylohyoid. Both muscles were considered to be important in the production of both lingual consonants and vowels. Based upon anatomical considerations and some of the studies reviewed above, it seemed likely that the genioglossus muscle should show a large amount of activity in the production of high front vowels (MacNeilage and Sholes 1964) and that the mylohyoid should be active in producing high vowels and alveolar and velar consonants (Faaborg-Andersen and Vennard 1964). In view of the fact that the genioglossus is a large muscle inserting into the anterior, medial, and posterior parts of the tongue, we decided to sample activity from at least two positions in this muscle.

Methods

Four subjects were used in this series of experiments. Each subject was a male, a speaker of general American English, and had perfectly normal speech. The subjects were seated in a sound-proof metal booth during the experiments. In sampling the emg signals from these muscles, bipolar electrodes made of thin (40 ga.) insulated magnet wire were used. This technique is described by Basmajian and Stecko (1962) and Hirano and Ohala (1969). The electrodes were inserted using 26 ga. 1 1/2" hypodermic needles. Emg activity during implantation of the electrodes was monitored visually on an oscilloscope, and auditorily over a loudspeaker. As long as the desired location for the electrodes has not been reached, the needle can be inserted further into the muscle tissues, while the placement of the electrodes is still being verified. The needle cannot be retracted, however, since the electrodes will remain behind.

The electrode insertion into the genioglossus muscle was made from below, through the supra-hyoid muscles. This approach avoids the problem, inherent in the oral approach to this muscle, of having the wires snag on the subject's teeth, or of their being accidentally pulled out by a vigorous movement of the lips or tip of the tongue. It also minimizes artefacts caused by movement of the tips of the electrode in the muscle.

In the genioglossus, the electrodes were inserted in two locations in 3 of the 4 subjects: (1) in the most superior fibers near the point of origin of the muscle, these being the fibers that insert into the anterior portion of the tongue, and (2) in the inferior fibers near the point of origin, that is the fibers that insert into the root or posterior portion of the tongue. The location of the electrodes was verified in two ways: by palpation of the tip of the needle being used to insert the electrodes, and by asking the subject to perform certain diagnostic gestures before the needle is withdrawn. We found that location of the electrodes in the genioglossus muscle can be determined by asking the subject to protrude his tongue. This action always resulted in strong emg activity if the electrodes were in any part of the genioglossus muscle, but little or no activity if the electrodes were in the nearby geniohyoid, mylohyoid, or digastric muscles. Different locations in the genioglossus were determined by having the subject make dental and velar stops. Stronger activity for dentals and weaker activity for velars indicated a correct placement in the anterior fibers. The opposite case, although often with no activity for dentals, indicated that the electrodes were correctly placed in the posterior genioglossus fibers. The method of insertion of the electrodes into the genioglossus is shown in Figure 2.5.

The mylohyoid was reached by inserting the wires into it from below, near the origin on the interior of the mandible, anterior to the hyoid bone. Electrode placement was determined as with the genioglossus, with the diagnostic gesture being a velar stop with a fixed jaw position. There are no nearby muscles which would be active under such circumstances. Mylohyoid insertion is shown in Figure 2.6. The technique used in implanting the electrodes and verifying their position is quite reliable.

Once the electrodes were correctly implanted, the subjects read a text involving the production of stressed CVC nonsense syllables in the frame "It's a ____ again". The frame was chosen to provide neutral tongue configurations on either side of the CVC syllable. There were two sets of CVC items. One used the consonants /p t d k g r/ and the vowels /i ɑ u/. In this set, all possible initial consonant plus vowel and all possible vowel plus final consonant combinations occurred, and for the consonants /t/ and /k/ all possible CVC combinations were included. The other set consisted of the vowels /i ι e ɛ æ ɑ ʌ ɔ o ɷ u/ in the syllable /p _ p/. Each of the resulting 36 items was read at least 20 times. The utterance types were randomized, the order being reversed for each subject, however the 20 tokens of each type were read consecutively, at a normal, relaxed rate of speech. The emg signals were amplified using a specially constructed IC differential preamplifier with a very high input impedance (Hubler 1967), and were recorded on 1/4" audio tape with a Viking Model 94, 4-channel recorder. The frequency response of the recorder was flat (± 3dB) from 50 to 7000 Hz. The audio signal from a condenser microphone was also recorded.

The emg signals were inspected for artefacts, and were averaged using a LINC-8 digital computer. The signals were reasonably clean, however they were high-pass filtered at 100 Hz before averaging to remove some 60 Hz hum and occasional low frequency movement artefacts. The computer averaging involves triggering an analog-to-digital conversion by the onset of the audio signal from an advanced head (500 msec advance at a tape speed of 7.5 ips) on the tape recorder, sampling the signal every 3.4 msec for 1.5 sec., full wave rectification, and summing of the signal present at each sample point. The resulting summed signals were then scaled and smoothed, and plotted graphically on a Hewlett-Packard Moseley X-Y Plotter. The plotting subroutine of the computer program incorporates a zero-signal reference mark at the end of each plotted line. This mark is not shown in the emg curves reproduced here, since they were trimmed for proper size in reproduction. The curves were plotted on mm-square graph paper, and all measurements of the averaged amplitude of the emg curves were made in terms of height of the curve in millimeters above the zero-signal reference mark. The audio signal from the standard playback head (not the advanced head) was processed in the same manner, except that it was not high-pass filtered before conversion. Further details on this technique are reported in Harshman and Ladefoged (1967). The scaling and smoothing factors were the same for each signal from each subject, however, they were different for different subjects. This is not important, since the absolute amplitudes of the signals, either raw or processed, cannot be compared across subjects. The absolute amplitudes of the emg signals from different muscles in one subject cannot be compared, either. However in both cases the relative variations in amplitude can be compared.

After the experimental session, the electrodes were removed quite easily and painlessly. It was found that the electrodes did not tend to slip during the course of an experiment, even though the experiments often lasted more than 2 hours. In several of the experiments, the first item of the text was repeated at the end of the experiment. No major differences between these control items and the first items in the text, in either the timing or amplitudes of the emg signals were noticed. At the conclusion of each experiment, a timing pulse was recorded on one channel of the tape recorder. This signal, usually a 4 Hz train of pulses 10 msec wide, was processed in a manner similar to the emg and audio signals, except that it was not filtered and was of course not averaged over 20 repetitions. This allowed an absolute time reference to be assigned to the plotted curves.

X-ray data

An X-ray film was also made of one subject (TS) reciting the entire text. The subject was seated comfortably in a chair, and his head was stabilized by a head rest. The film was made by photographing the output of a Standard X-Ray image intensifier tube using a 16 mm. Auricon camera, at a filming rate of 24 frames-per-second. The sound track was recorded optically on the film and magnetically on a separate tape recorder.

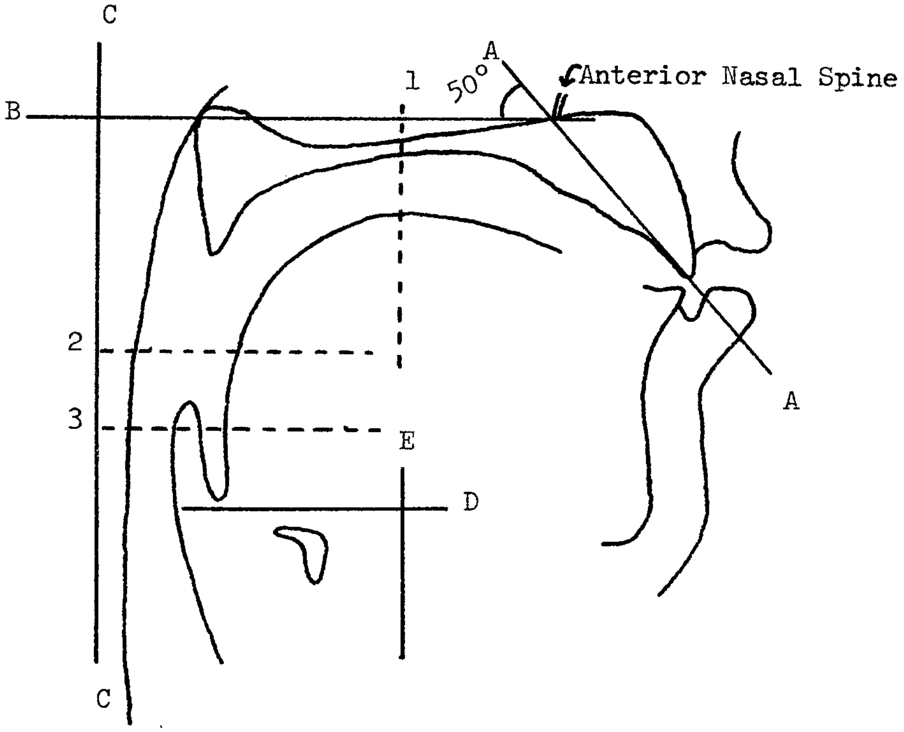


Figure 2.7

One frame of the X-ray film showing the coordinate system for measuring the tracings superposed on the tracing.

The X-ray film was taken on the same day that the emg data were collected for this subject. The electrodes were left in place during the filming, however they could not be seen on the film due to their small size relative to the resolving power of the X-ray system. Further details concerning the X-ray apparatus are reported in Delattre (1967).

The X-ray film provided an image which allowed the outline of the lips, soft and hard palates, and the tongue to be visualized and traced. Hard structures posterior to the pharyngeal wall could not be clearly identified. In this subject, however, there were two clear reference points in the fixed structure of the head which permitted the construction of a set of measuring coordinates: the tip of central maxillary incisors and the lower part of the anterior nasal spine. The film was analyzed in the following way. Certain key utterances were chosen for tracing. In this selection, we were most interested in the behavior of the tongue and hyoid bone in the production of tense and lax vowels in the labial environment (/pip/, /pep/, /pɩp/, /pep/, /pɑp/, and /pup/), in the differences in vocal tract shape for the voiced and voiceless dental stops, and in the production of the velar voiceless stop /k/. The film was examined in a sound editor, and the film frame indicating a dental closure for the /ts/ of "It's" at the beginning of the frame for each selected CVC word was marked. The film was transferred to a PDL Percepto-Scope still-frame projector, and each of 15 frames following the marked frame was traced. The CVC syllable was always included in these 15 frames. Each tracing was done as illustrated in Figure 2.7. The position of the anterior nasal spine and the tip of the central maxillary incisors was carefully marked. The tracings were done lifesize. All tracing was done by one person. The tracings were carefully examined, and the distance between the lower part of the anterior nasal spine and the tip of the incisors was measured. The average error in this measurement was typically less than 1.5 millimeters. Apart from the pincushion distortion inherent in this type of cinefluorography, the X-ray data reported here should be considered to have an uncertainty of <2 millimeters.

A coordinate system was constructed to enable measurement of the tracings. The coordinate system is shown superimposed on a sample tracing in Figure 2.7. A line (A) was drawn between point N (the lower part of the anterior nasal spine) and the tip of the upper central incisors. A line (B) was constructed passing through point N. Six cm. posterior to N, a line (C) was constructed perpendicular to B. The angle between A and B (50°) was chosen so that lines B and C would be approximately parallel to the maxilla and posterior pharyngeal wall respectively. Measurements of the tongue position were made along lines 1, 2, and 3. The position of the upper anterior point of the hyoid bone was measured with reference to lines D and E (parallel to B and C). Line 1 is indicative of the tongue position for high vowels and velar stops, although it should be noted that this line is somewhat anterior to the central point of the vocal tract occlusion for the palato-velar [k͔] in /kik/. Measurements 2 and 3 indicate the position of the posterior third of the tongue, and are considered to be indicative of the action of the posterior genioglossus muscle. The vertical displacement of the hyoid bone is indicative, within a certain range, of the length of the mylohyoid muscle. There is little reason to suppose that there is a linear relation between vertical hyoid displacement and mylohyoid length, since the mylohyoid does not exert an upward pull on the hyoid bone parallel to the film plane. Nevertheless it is felt that this measurement is the best available indication of mylohyoid activity in the X-ray tracings.

Results

In discussing the results of this experiment, we shall at present concentrate on the role played by the muscles in deforming the tongue for characteristic speech gestures. For this reason, the results will be presented under headings of the name of the muscle. In Chapters 3 and 4, the relation of these results to models of speech production and to the formulation of a set of phonological features will be discussed.

In general, we found that there were no consistent and reliable differences in the emg activity in these muscles of the tongue between voiced and voiceless consonants made at the same place of articulation. There was, of course, no activity for the labial stop /p/. For these reasons, we shall discuss mainly the gestures associated with the voiceless alveolar stop /t/ and the voiceless velar stop /k/.

1.) Mylohyoid

This muscle shows a very consistent pattern of activity. It is markedly active in the gesture producing a velar stop, and shows lesser activity for alveolar stops. In three of the four subjects, the mylohyoid appears to be most active for a /k/ in the environment of /u/ and /a/, least active for /t/ before /i/. The velar closure before a back vowel is retracted, the closure before a high front vowel is advanced. Thus, in these 3 subjects, the mylohyoid shows a consistent pattern: [k͕], [k͔], [t] in decreasing order of amount of activity (see Figure 2.8). In one subject (GA), there was more mylohyoid activity for alveolar stops than for any velar or palato-velar stop.

By contrast, the mylohyoid shows little activity for vowels. Where there is activity, it tends to occur for front vowels rather than for back vowels, although it is not consistently greater for high vowels than for low vowels. Table 2.1 shows the data for 3 subjects for all 11 vowels in the environment /p _ p/. The units of amplitude are arbitrary, and

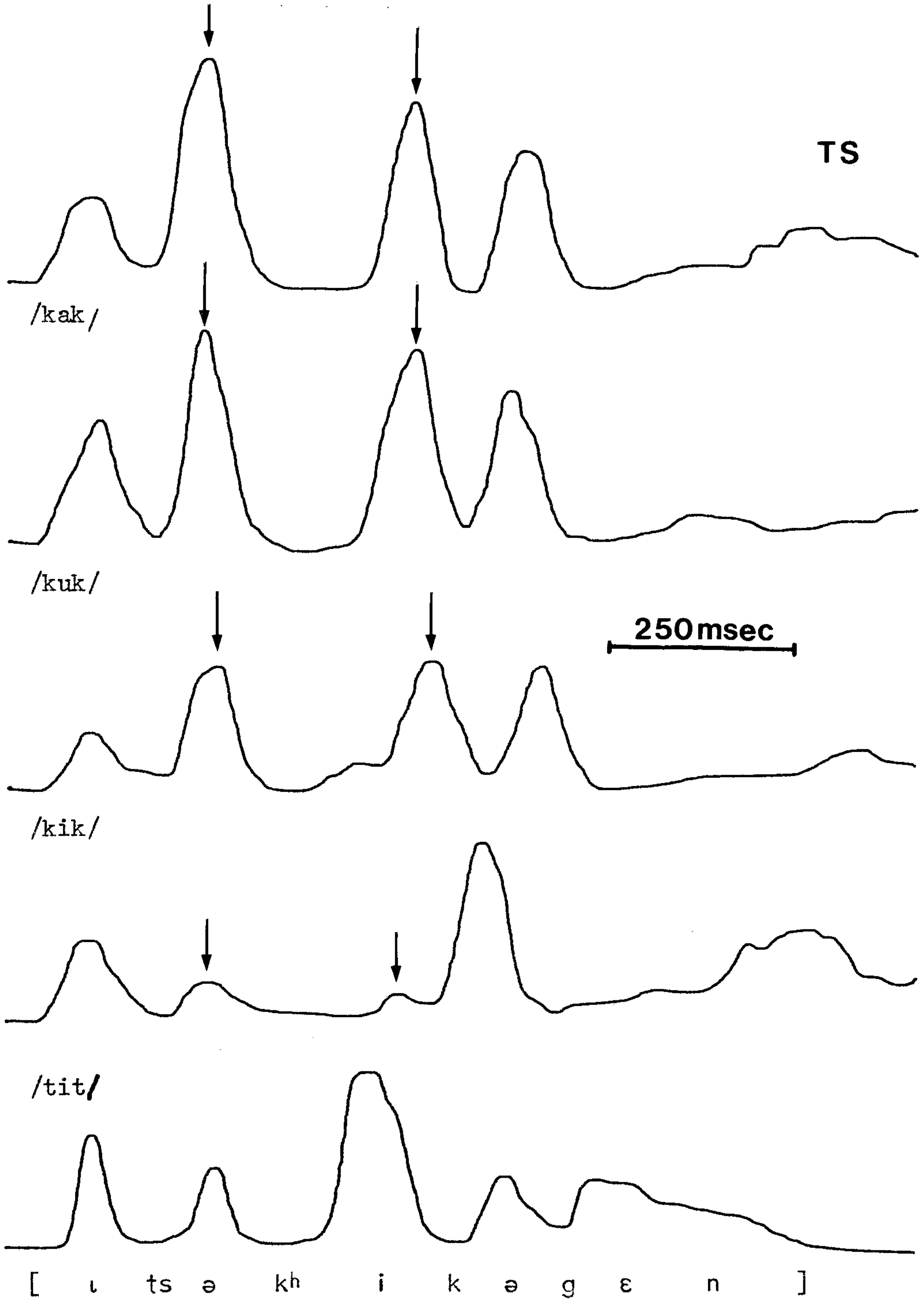


Figure 2.8

Averaged emg signals from the mylohyoid muscle for the items //kak/, /kuk/, /kik/, and /tit/. The audio signal is for the item "It's a /kik/ again." Subject: TS. The arrows above the averaged curves indicate the peaks for the initial and final consonants in the CVC word.

	Subjects		
	TS	GP	GA
/i/	7	4	6
/I/	6	6	-
/e/	5	3	9
/E/	6	7	6
/æ/	4	5	-
/ʌ/	-	-	-
/ɑ/	-	-	-
/ɔ/	3	6	-
/o/	-	-	-
/U/	3	-	4
/u/	5	-	-

Table 2.1

are not, as mentioned above, comparable across subjects. Data for the fourth subject, JO, is not available for the set of items involving the 11 vowels, due to a malfunction of the recording apparatus. In general, the mylohyoid muscle is not markedly active in vowel production. There is a certain amount of activity for front vowels, however this may represent only a small degree of stabilizing activity for tongue position. This finding somewhat contradicts the results reported by Faaborg-Andersen and Vennard (1964) mentioned above. However they were dealing with steady-state, sung vowels, which are typically different in manner of production from vowels in a normal utterance. They did not investigate mylohyoid muscle activity in consonants.

2.) Genioglossus, anterior fibers

The activity of the anterior and posterior fibers of the genioglossus was similar in some respects, however it also differed in many ways. In Figure 2.9, the activity of these two portions of the genioglossus is plotted,

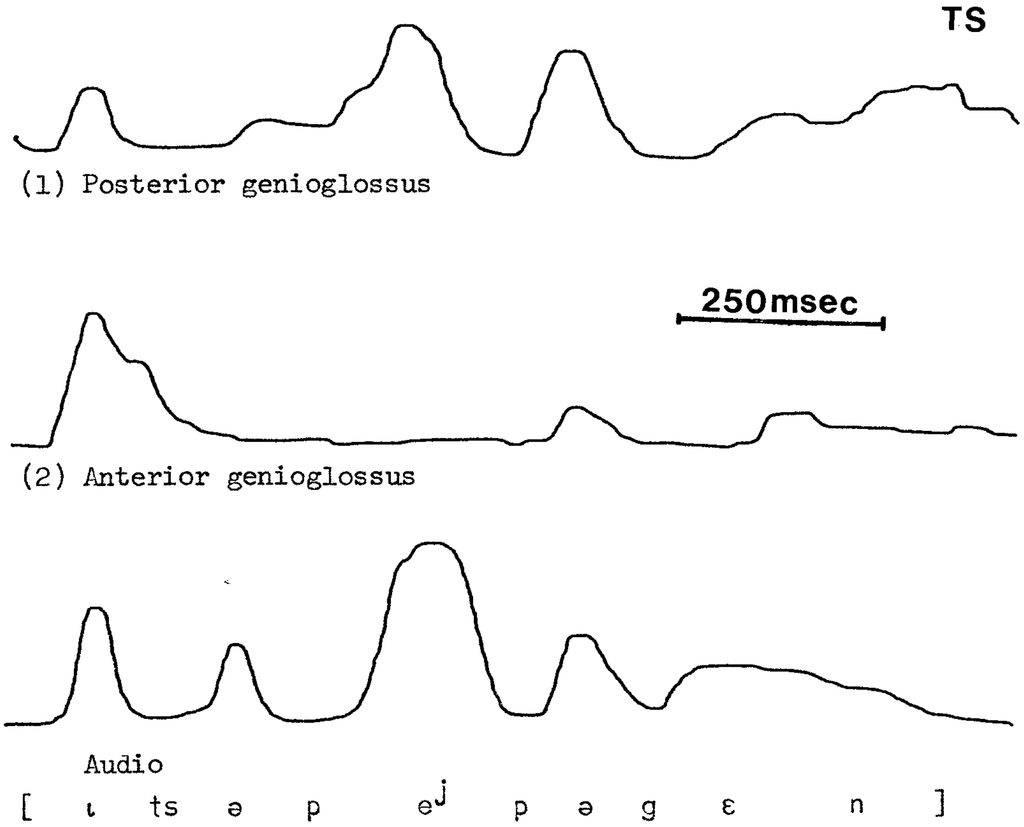


Figure 2.9

Averaged emg signals from (1) the anterior fibers of the genioglossus muscle and (2) the posterior fibers. The sentence is "It's a /pep/ again."

along with the audio signal, for the utterance "It's a [pejp] again." Notice that the posterior genioglossus is quite active during the vocalic nucleus of [pejp]. This activity corresponds to the high front tongue position for the glide component of the nucleus. The anterior portion of the genioglossus, however, is totally inactive at this time, since contraction of this portion of the muscle would be antagonistic to the gesture fronting and raising the tongue body.

The anterior genioglossus fibers showed the most activity for the alveolar stops /t/ and /d/. It also showed some activity for the /n/ in the frame. The peaks of this activity occurred before the stop closure, demonstrating that this is closure-related activity, rather than activity associated with the stop release. It is interesting to note that this activity varies in degree depending upon the phonetic context. Table 2.2 demonstrates the emg amplitude (in arbitrary units) for the /t/ in all possible positions from the anterior genioglossus (as well as the mylohyoid and posterior genioglossus). The amplitude of the initial stops varies slightly, being somewhat higher for the /t/ before /u/ than before /i/ or /ɑ/. In the stops in final position in the syllable, however, the activity

S: TS	Ant. Genioglossus		Post Genioglossus		Mylohyoid	
	C_1	C_2	C_1	C_2	C_1	C_2
tat	43	43	20	27	16	17
tit	44	30	26	-	12	8
tut	51	20	19	-	11	12
kat	5	20	22	17	58	27
kit	9	10	28	-	51	6
kut	18	16	26	-	51	22
tak	33	22	18	24	11	22
tik	50	20	27	52	13	54
tuk	40	24	27	45	7	34

Table 2.2

Amplitude (arbitrary units) of the Averaged emg peaks for /t/ in all positions. Subject: T.S.

for the /t/ following /a/ is considerably greater than for the /t/ following either /i/ or /u/. This is shown in Figure 2.10. Since the gesture for /ɑt/ requires that the tongue be returned from a low position compared with /it/ or /ut/, this is not very surprising.

At least in this one subject, the anterior portion of the genioglossus appears to be most active in positioning the tongue near the alveolar ridge for alveolar stops. The peaks of the emg activity for /t/ occur an average of 60 msec before the stop closure (range 50-75 msec). The actual closure gesture for the alveolar consonants is probably made by the intrinsic muscles, principally the superior longitudinals and the transverse muscle.

The activity of the anterior genioglossus in the production of vowels is fairly weak and diffuse. We had anticipated this result, since the anterior portion of the tongue does not take on definite shapes in the production of vowels.

3.) <u>Posterior genioglossus</u>

In general, the posterior genioglossus showed the highest activity during velar consonants and tense, high vowels. In two of the subjects (TS and GA) there was activity for syllable-initial alveolar stops, although this was always considerably less than for the velars. Order rankings for the activity of this muscle for various consonants are shown in Table 2.3. The activity of this muscle for velar stops is shown in Figure 2.11.

Subjects

	TS	GP	GA	JO
k/ - ik	1	1	1	1
k/ - ak	3	2	2	2
k/ - uk	2	3	3	3
k/ ki -	2	1	1	1
k/ ka -	3	2	3	2
k/ ku -	1	3	2	3

Table 2.3

Order rankings of the peak amplitude for /k/ in word initial and word final position: Posterior genioglossus.

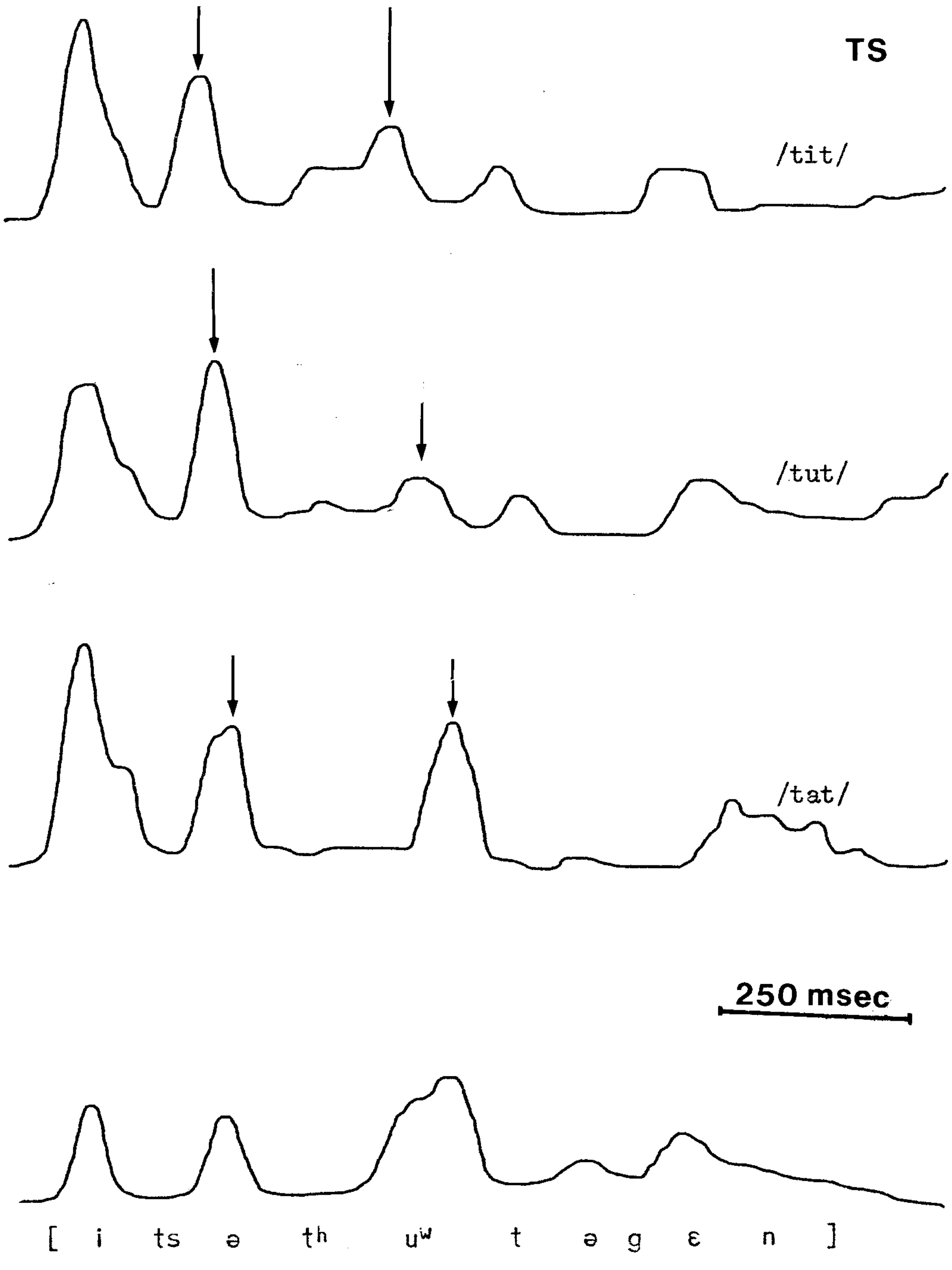


Figure 2.10

Averaged emg signals from the anterior genioglossus muscle for /t/ in various vocalic contexts. The audio signal is "It's a /tut/ again." The arrows above each curve point out the peaks in the emg signal associated with the initial and final consonants in the CVC word.

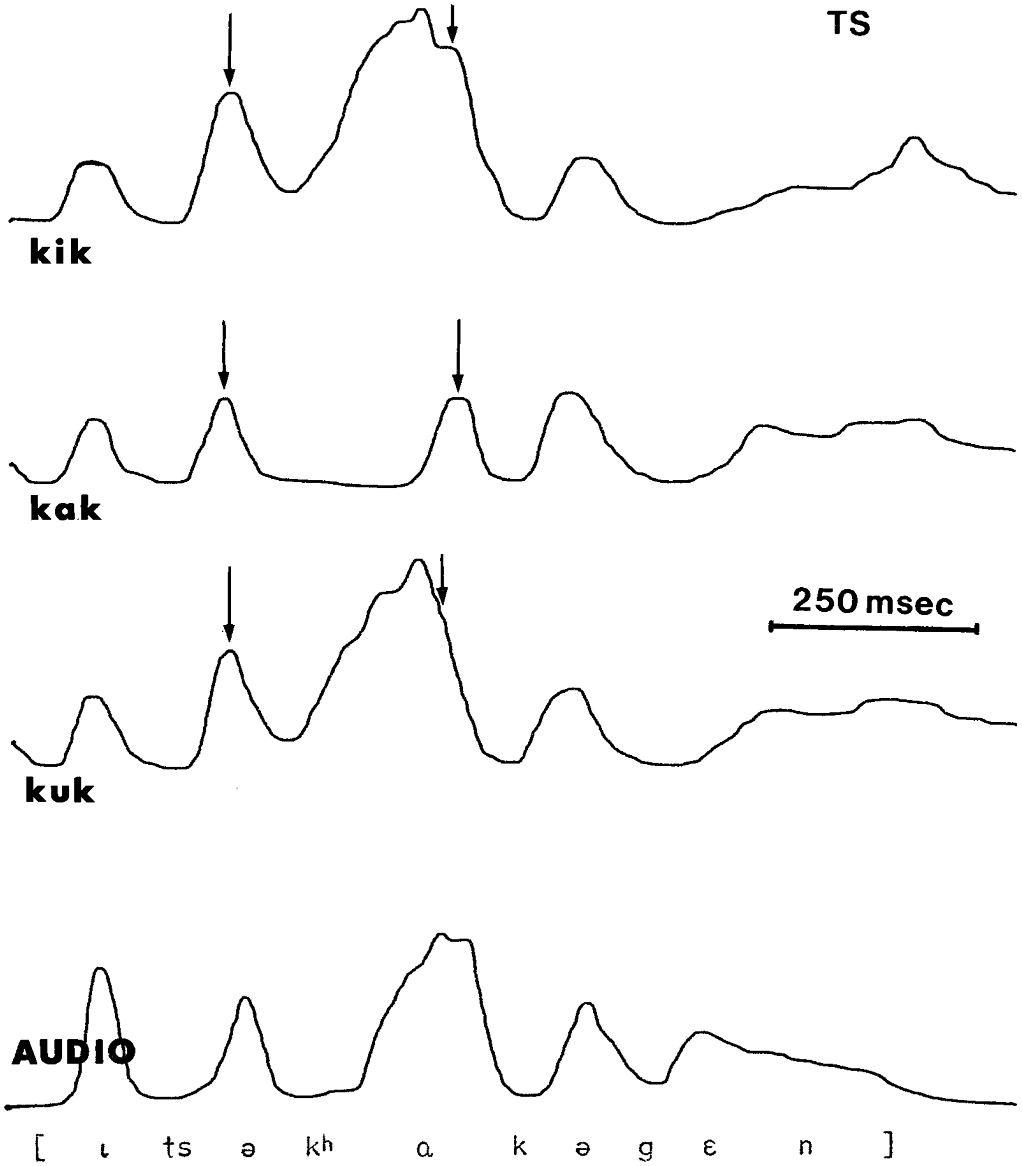


Figure 2.11

Averaged emg signals from the posterior genioglossus muscle for /k/ in various vocalic contexts. The audio signal is "It's a /kak/ again." The arrows above each curve represent the peaks in the emg signal associated with the initial and final consonants in the CVC word.

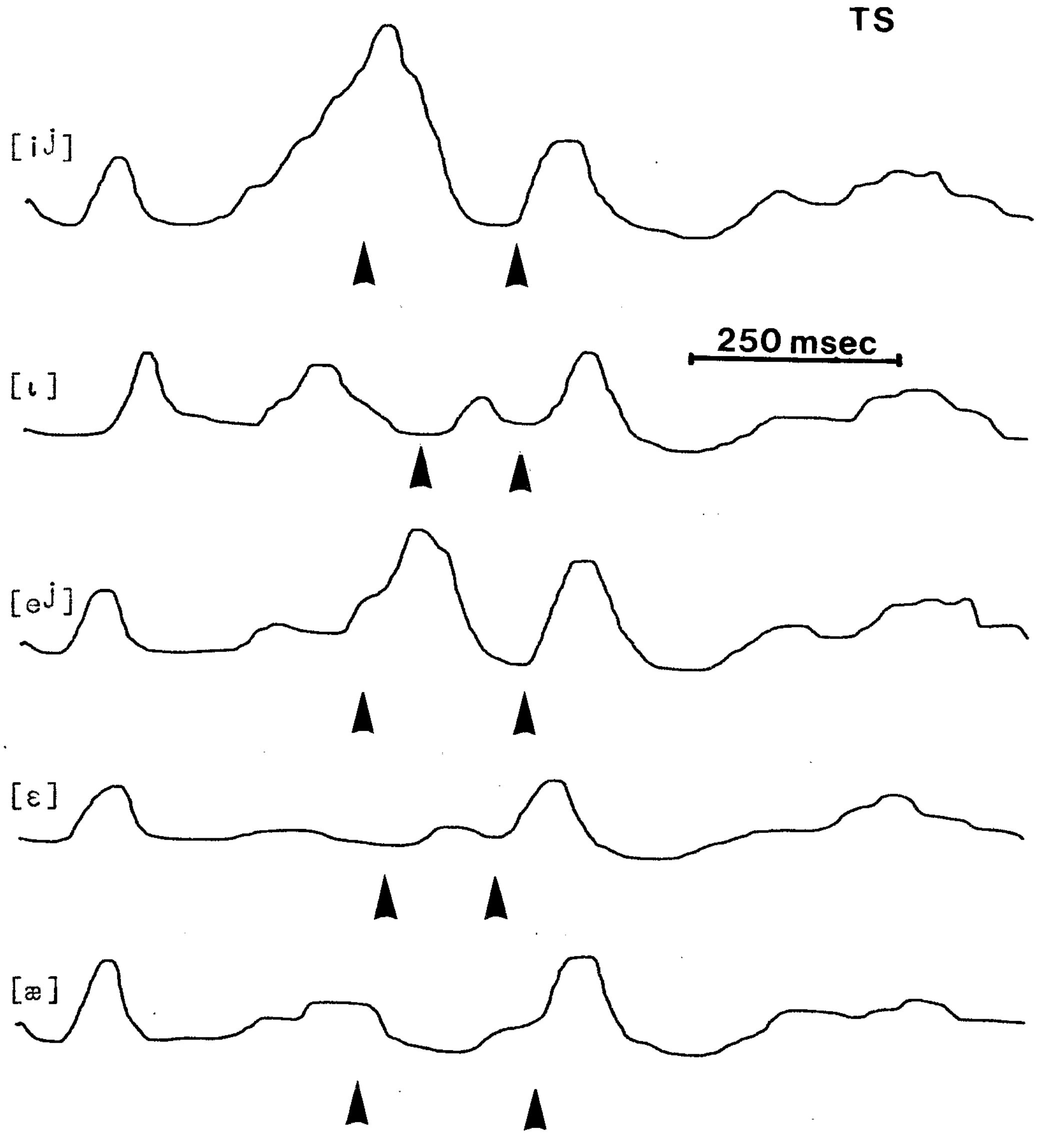


Figure 2.12a

Activity of the posterior genioglossus muscle for front vowel nuclei in the environment "It's a /p___p/ again." The arrows below each curve indicate the onset and offset of voicing for the vowel nucleus.

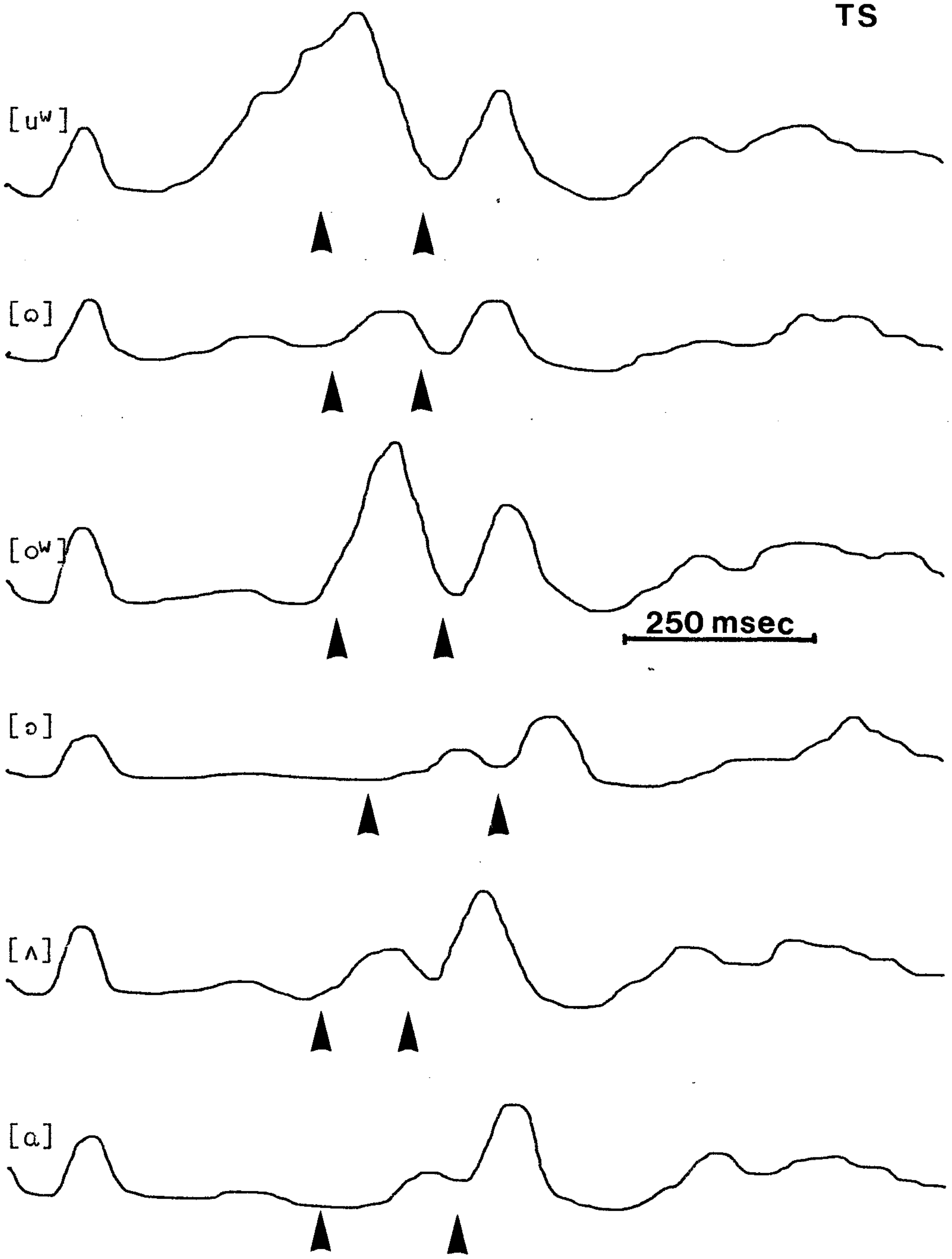


Figure 2.12b

Averaged activity of the posterior genioglossus muscle for back vowel nuclei in the environment "It's a /p___p/ again." The arrows below each curve indicate the onset and offset of of voicing for the vowel nucleus.

The most notable aspect of the posterior genioglossus activity is the large amount of activity manifested in the production of tense, high vowels. In Figure 2.12, the activity of this muscle for the 11 vowels in the labial environment for one subject (TS) is plotted. The first arrow on each emg curve indicates the onset of voicing for the vowel in that CVC syllable, the second arrow is the voicing offset. Note that there is little or no activity for the low vowels and the mid non-tense vowels. With the vowels /ɛ/, /ʌ/, /a/, /ɔ/, and /ɷ/ there is a small peak of activity, however this occurs just before the offset of voicing. This activity is clearly produced by a gesture returning the tongue to a more neutral configuration in the vocal tract. If this is the case, however, it is difficult to explain why this peak should be so large at the offset of the stressed vowel /ʌ/ which is, presumably, not radically different from the reduced /ə/ of "again", the next vowel in the sequence. These findings confirm the claim of MacNeilage and Sholes (1964) that the activity of their posterior surface electrodes during vowels with a high front tongue position is due to contraction of the posterior fibers of the genioglossus. They did not see activity during the high back vowel /u/, which was found in this experiment, however.

One aspect of this activity which is worth questioning is whether this activity is due to the positioning of the tongue for the vowel itself, or to the final glide which is present in each of the tense, nonlow vowels in the pronunciation of all subjects. There are several facts which suggest that the large peak in each curve is caused by the glide component of these vowels. In a CVC utterance where there is emg activity for some isolated muscle associated with the vowel, and none with the consonant, as is the case here, the peak of the integrated or averaged emg will usually precede or be coincident with the onset of voicing for the vowel, as shown in Figure 2.13.

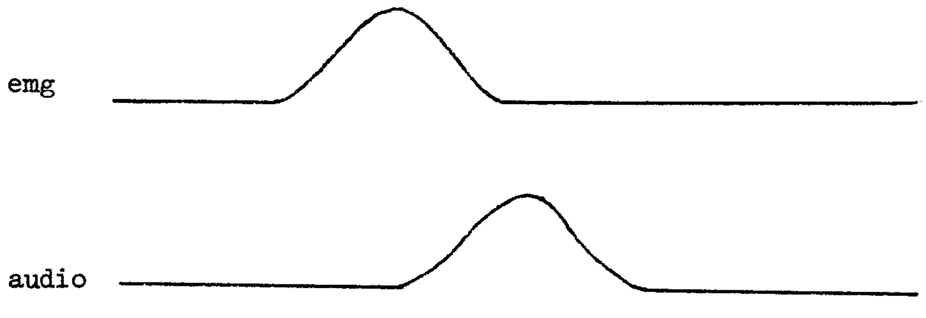


Figure 2.13

In these vowels, the peaks for /e/ and /u/ occur well after the vowel onset, and the peak for /i/ occurs shortly after onset. Also, the average emg curves for these vowels are slightly skewed to the right. The curve for /u/ in Figure 2.12 demonstrates this quite strikingly. There are two, perhaps three peaks in that part of the emg signal for the vowel. The first is in the location expected for the production of the vowel shape itself. The later peaks can only be attributed to the velar part of the labio-velar glide following the /u/, i.e. to the [w]-component of [uw]. The same holds true for the vocalic nuclei [ij] and [ej]. It is only in the production of the glides that the posterior genioglossus activity begins to approach the levels found in the velar stops (Figure 2.11).

One other interesting aspect of the behavior of the posterior genioglossus is the patterns of activity for this muscle in the production of voiced and voiceless alveolar stops. The activity of the posterior genioglossus in all subjects was almost identical for the stops /t/ and /d/. Figure 2.14 shows this activity for one subject (TS). In addition, the mylohyoid activity for these stops was also quite similar. If anything, it is stronger for the /t/ than for /d/.

The behavior of the posterior portion of the genioglossus muscle in speech can be summarized as follows: this muscle is used to pull the root of the tongue forward, away from the posterior pharyngeal wall. This gesture is a major component in the articulation of front and back high vowels /i/ and /u/. The muscle is more active,however, in non-obstruent sounds involving a greater degree of closure in the vocal tract, i.e. glides, than it is in high vowels. It is not simply the case that the greater the closure in the vocal tract, the greater the activity of the posterior portion of the genioglossus. The activity for high or mid, front vocalic nuclei like [ij], [uw], [uw], or [ej] is greater than for velar consonants, as can be seen clearly in Figure 2.11. In the case of the velar stops, the actual closure of the vocal tract is made principally by the activity of the mylohyoid muscle. The genioglossus appears to act to "bunch" the tongue into a position suitable for high vowels *and* for velar stops. For the vowels,the mylohyoid is minimally active, but for the stops there is a large burst of activity in this muscle which serves to lift the body of the tongue and thus make the closure. The activity patterns of these muscles appear to be finely graded, the amount of activity present for any particular phoneme being a function of the immediate phonetic context. The context effects present in this data will be discussed in detail in the following chapter.

EMG EXPERIMENT # 2

The data presented above give a reasonable picture of the activity of those extrinsic tongue and suprahyoid muscles which can pull the tongue, or selected portions of the tongue, forward in the oral cavity, thus deforming it for the articulation of certain consonants and vowels, as well as raising the body of the tongue for stop consonants. These data, however,

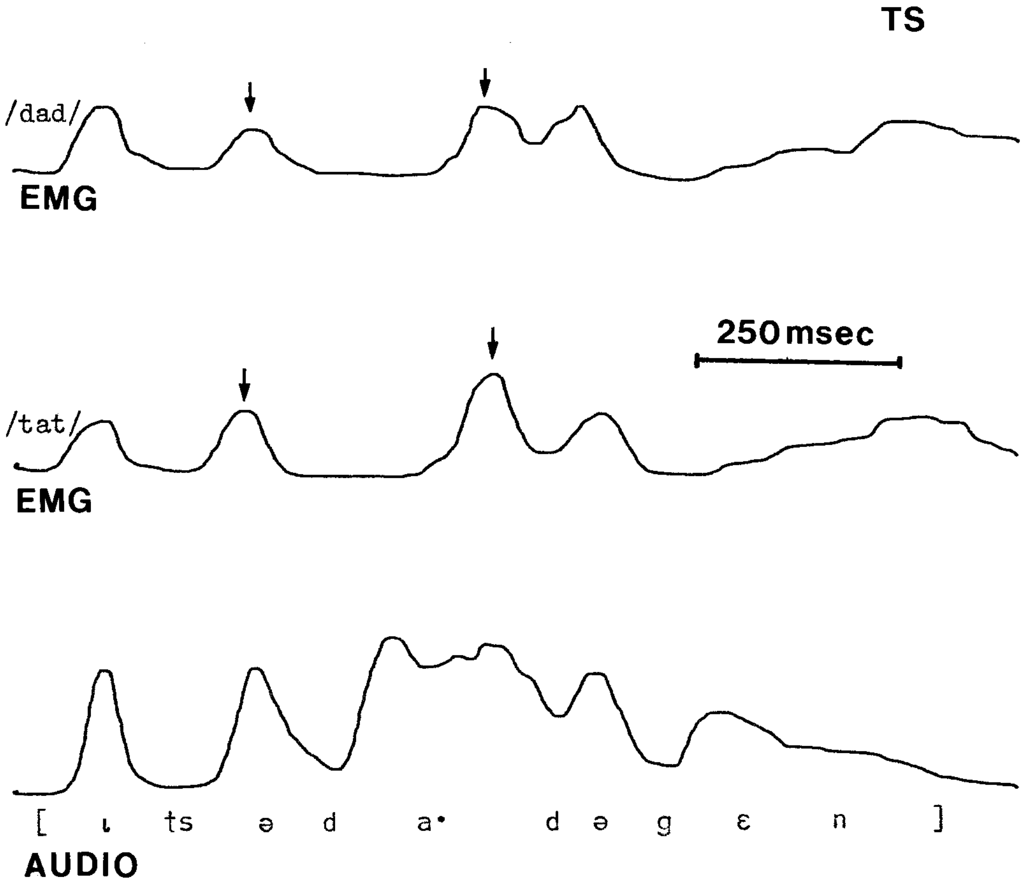


Figure 2.14

Averaged emg signals from the posterior genioglossus muscle during the utterances "It's a /dad/ again," and "It's a /tat/ again." The audio signal is for /dad/. The arrows point to the peaks in the emg curves for the initial and final consonants in the CVC words.

provide no information concerning the control of the tongue in producing low front and back vowels, and an incomplete picture of the production of high back vowels and velar consonants. Of the remaining extrinsic tongue and suprahyoid muscles, the most important are the hyoglossus, the styloglossus, the glossopharyngeus, and the stylohyoid. The palatoglossus muscle was investigated by Fritzell (1969), who used an oral approach and bipolar wire electrodes. He found that the muscle was active for velars [ŋ] and [g] in the VCV syllables [əgV] and [əŋV], where the vowel was [i], [ɑ], or [u]. There was considerably more activity for the [ŋ] than the [g], showing that the muscle participates in an active lowering of the velum. It seems likely, considering the size of this muscle and the relative sizes of the velum and tongue, that the palatoglossus contributes more to movement of the velum than it does to tongue movement.

The remainder of these muscles are much more difficult to isolate and record from than the mylohyoid, palatoglossus, or genioglossus. With the possible exception of the hyoglossus, they are all difficult to approach transcutaneously. The styloglossus near its point of origin is a deep muscle, and lies near important arteries (the internal carotid) and nerves (the vagus). The stylohyoid is also difficult to approach, for the same reasons. Similar problems would be found in trying to insert electrodes into the glossopharyngeus transcutaneously.

An oral approach to these muscles is possible; however, there are complications. The presence of the fine electrode wires in the mouth does not impede normal speech. Nevertheless, slight pulls on the wires can cause large movement artefacts in the recorded signals. More importantly, the oral approach presents problems in that it is difficult to direct the electrode implantation carefully, since the subject cannot usually inhibit his gag reflex sufficiently. Palpation of the tip of the needle cannot be done, and it is difficult to hold the needle in place for a long enough time to allow monitoring of the emg signals before withdrawing the needle, leaving the electrodes implanted. Nevertheless, it was felt that a knowledge of the anatomy of the area would give some chance of success to an "empirical approach."

Methods

In this experiment only one subject was used. The subject is a speaker of a variety of British English. His vowels are rather different auditorily from the American English vowels of the subjects in the previous experiment, nonetheless the basic gestures employed are probably rather similar. The consonants are quite similar to those of American English. The experiment was carried out in the minor operating room of the Head and Neck Surgery Department, UCLA Health Sciences Center. The subject was prepared for the experiment by applying a topical anesthetic to the mucuous membrane of the tongue and oral pharynx, and injecting a 1% Xylocaine solution in the immediate area of insertion. These precautions were necessary to inhibit the gag reflex, and to alleviate any pain upon insertion of the electrodes.

Bipolar fine wire electrodes were used, as in the previous experiment. These were implanted with 27 ga. 1/2" hypodermic needles. The signals were preamplified as before, and were recorded on a 7-channel Ampex SP-300 tape recorder, using direct recording (no equalization). The frequency response of the recorder is flat (± 3 dB) from 30 Hz to over 30 kHz. The audio signal was recorded in the same manner, using a Sennheiser MD21 dynamic microphone, placed about 6" from the subject's lips.

Electrodes were inserted in 3 locations in this subject:

A -- into the side of the oral cavity, at the base of the tongue, just below the left palatine tonsil (intact in this subject), the direction being anterior and slightly caudal, to a depth of about 3/8 - 1/2". It was felt that this location would record mainly from the styloglossus muscle.

B -- on the anterior portion of the lateral pharyngeal wall, about 1-2 cm inferior to location A, same depth and general direction. This location might pick up either styloglossus or glossopharyngeal activity.

C -- on the left side of the posterior pharyngeal wall, just above the level of the tip of the epiglottis. It was felt that this location would record activity from the glossopharyngeal fibers of the superior pharyngeal constrictor.

After the electrodes had been inserted, the subject was placed in a comfortable sitting position. Within a half-hour after the anesthesia, the subject's speech sounded perfectly normal. The electrode wires did not cause him any discomfort. He then performed the following gestures:

1.) Pronouncing a voiced pharyngeal fricative ([ʕ])

2.) Protruding the tongue

3.) Opening the jaw

4.) Making a velar closure ([aka])

5.) Raising the larynx

6.) Depressing the larynx

7.) Clenching the jaw

8.) Swallowing

For the pharyngeal fricative activity was noted at B and C, very little at A. Protruding the tongue showed some activity at all locations. Opening the jaw showed no significant activity at any electrode location. When the subject made a velar stop, there was activity for location A, some for B and almost none for C. No large emg signals were produced at any location for raising and lowering the larynx, or for clenching the jaw. In swallowing, there was a small amount of activity at location A, slightly less than for a velar stop, but large amounts of activity at location B and C. The activity at A reached its maximum earlier than at locations B and C. The emg signals for the swallow are shown in Figure 2.15. These patterns of activity were precisely those expected by our initial hypotheses. The electrodes in location A record activity for gestures involving a retraction plus elevation of the tongue body, as in making a velar stop, or in one of the stages of swallowing. This location did not show any activity in low vowels. Figure 2.16 shows one token of the syllable [aka]. In location B, the electrodes produced a signal having some of the characteristics of both location A and C during speech activities, although it differed during certain non-speech activities, such as swallowing. Location C was most active in swallowing, and in pharyngeal constriction. Figure 2.17 shows the emg activity for all 3 locations during the pharyngeal fricative [ʕ].

Based upon these results, we will assume that location A represents emg activity of the styloglossus muscle, and location C represents activity of the glossopharyngeal part of the superior pharyngeal constrictor muscle. The emg signals recorded from location B were quite varied, and could not be positively identified. They appear to reflect the activity of more than one muscle. It should be emphasized that these identifications are tentative. Much more exact work, of the sort demonstrated in Fritzell (1969), will be necessary to map the precise function of these muscles and to determine replicable procedures for inserting electrodes into them.

The subject next read a text composed of the words *peeve, fob, kipper, pave, poop, cough, keg, cog, keeper, cooker, paw, pipper, coop,* and *cope*. Each word was pronounced 20 times in the frame "It's a ____." At the end of this text, the subject produced 10 tokens each of the syllables [bi], [ba], [bɑ], [bɔ], [bo], and [bu]. The vowels in these CV syllables had approximately their cardinal values. The signals were computer processed in a manner similar to those in the first experiment. Each signal was high pass filtered at 120 Hz, and averaged using the AVG 2 program run on a PDP-12 computer. The signals were sampled at a rate of 330 Hz, or one sample every 3 msec. In the initial examination of the raw data, it was found that most of the tokens of each utterance contained movement artefacts. These were apparently caused by slight pulls on the wire during stop closures. High pass filtering at 120 Hz did not remove them completely, since most of them involved a rapid base line shift having high-frequency components. Since filtering at a higher cut-off frequency would have attenuated the signal excessively, they could not be completely eliminated. They can, however, be easily identified in the averaged curves

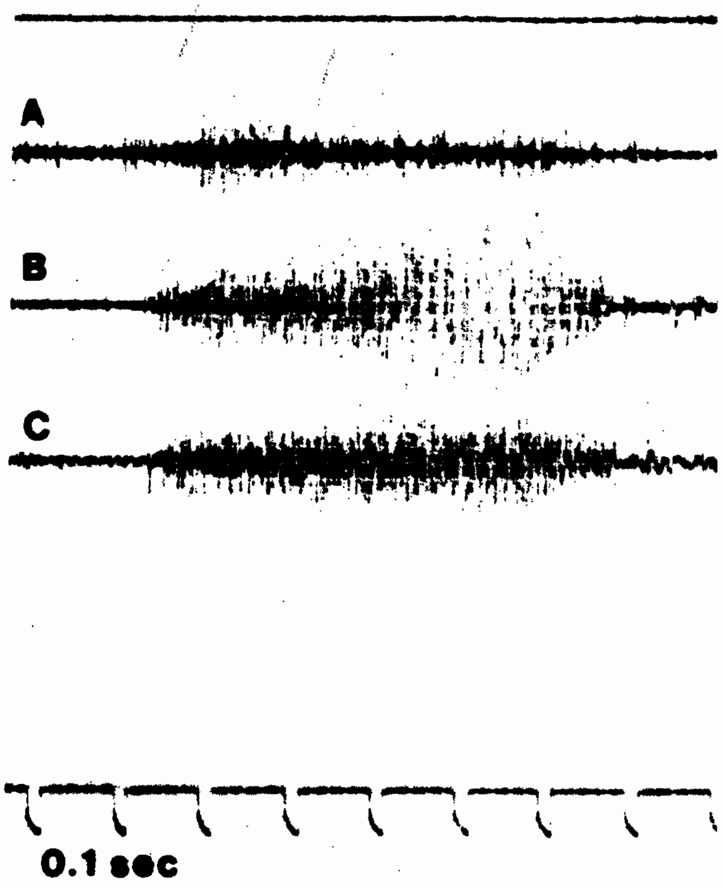


Figure 2.15

Emg from electrode locations A, B, and C (experiment #2) during a swallow.

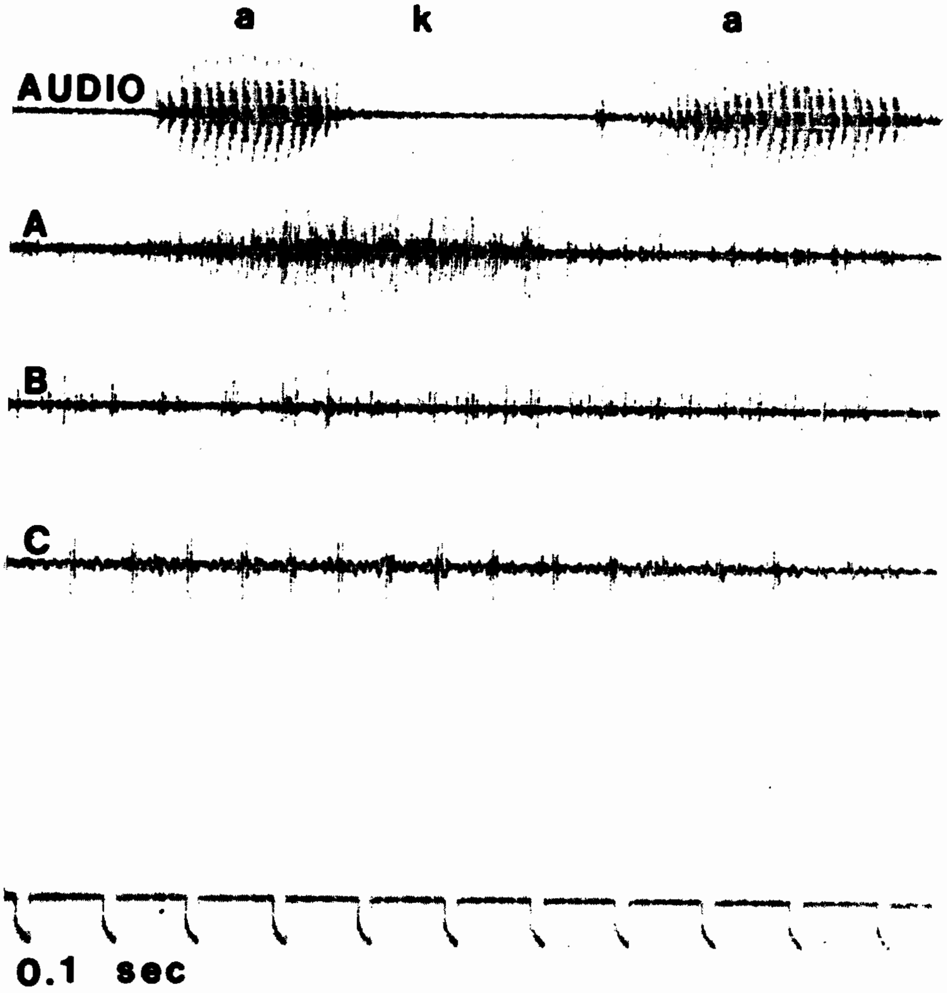


Figure 2.16

Emg from electrode locations A, B, and C (experiment #2) during one token of the utterance /aka/. The top trace is the microphone signal.

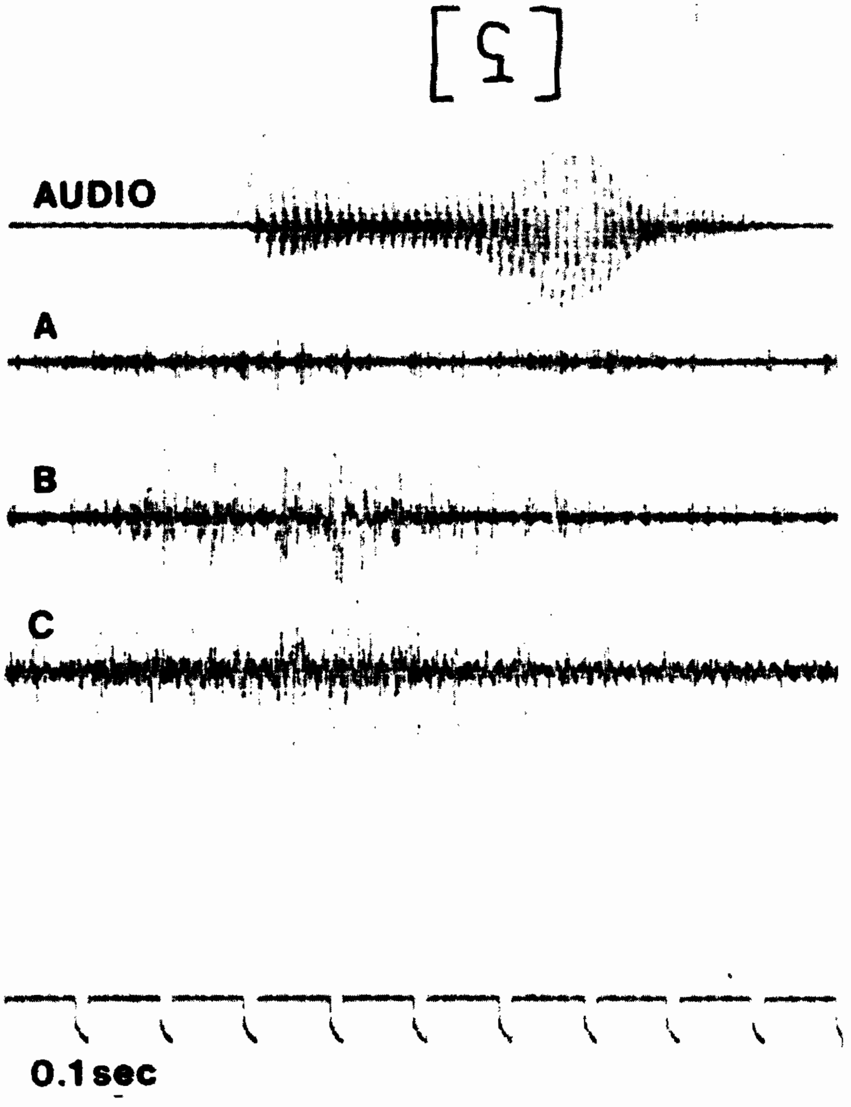


Figure 2.17

Emg from electrode locations A, B, and C (experiment #2) during a voiced pharyngeal fricative [ʕ]. The top trace is the microphone signal.

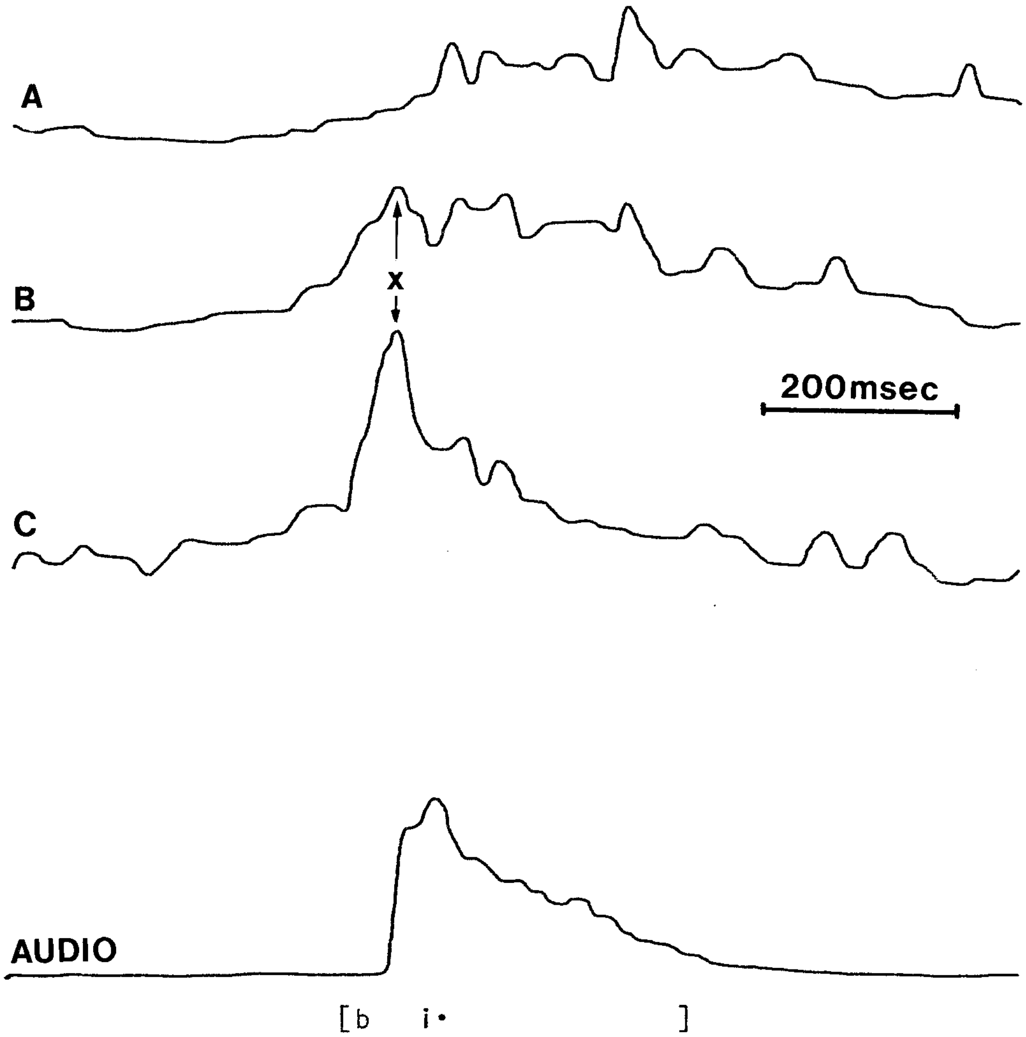


Figure 2.18a

Averaged emg signals from the electrode locations A, B, and C of experiment #2 for the CV syllable [bi:]. The X's indicate spurious peaks in the emg averaged curves which were produced by movement artefacts.

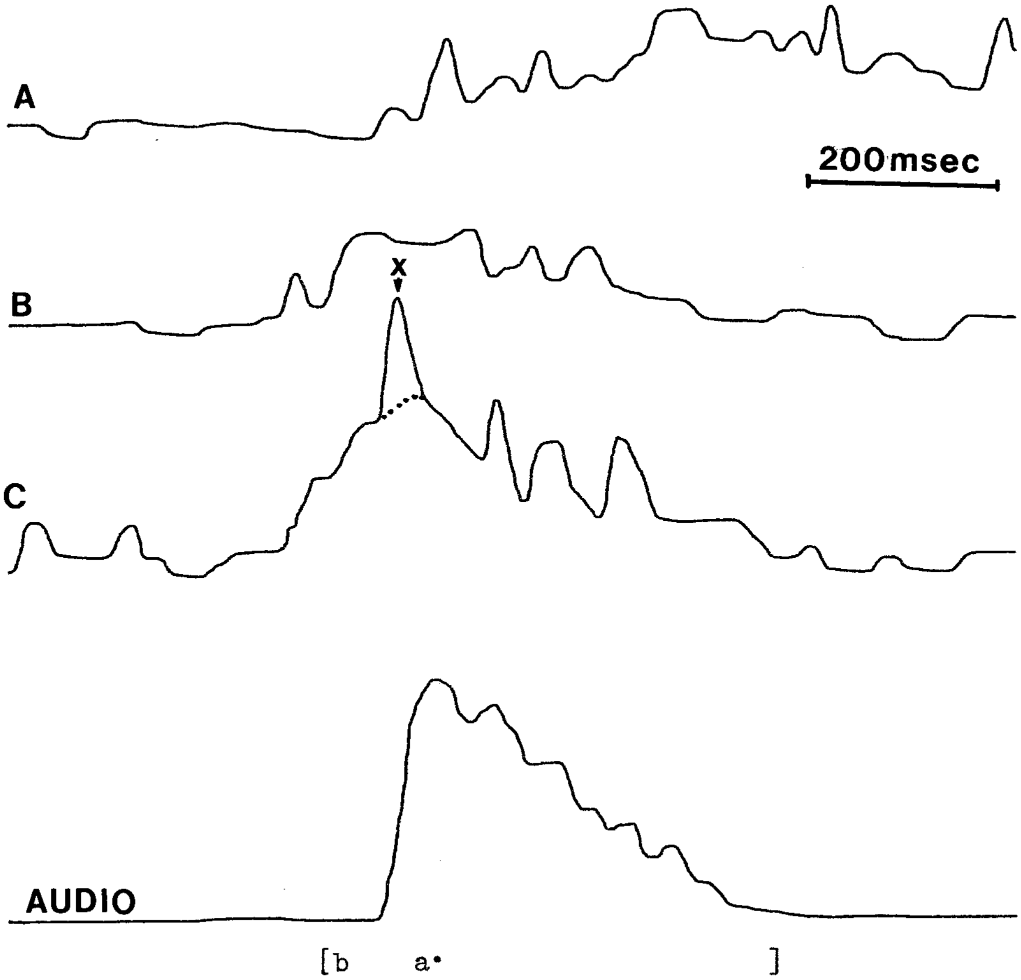


Figure 2.18b

Averaged emg curves from the electrode locations A, B, and C of experiment #2 for the CV syllable [ba:]. The X's indicate spurious peaks produced by movement artefacts.

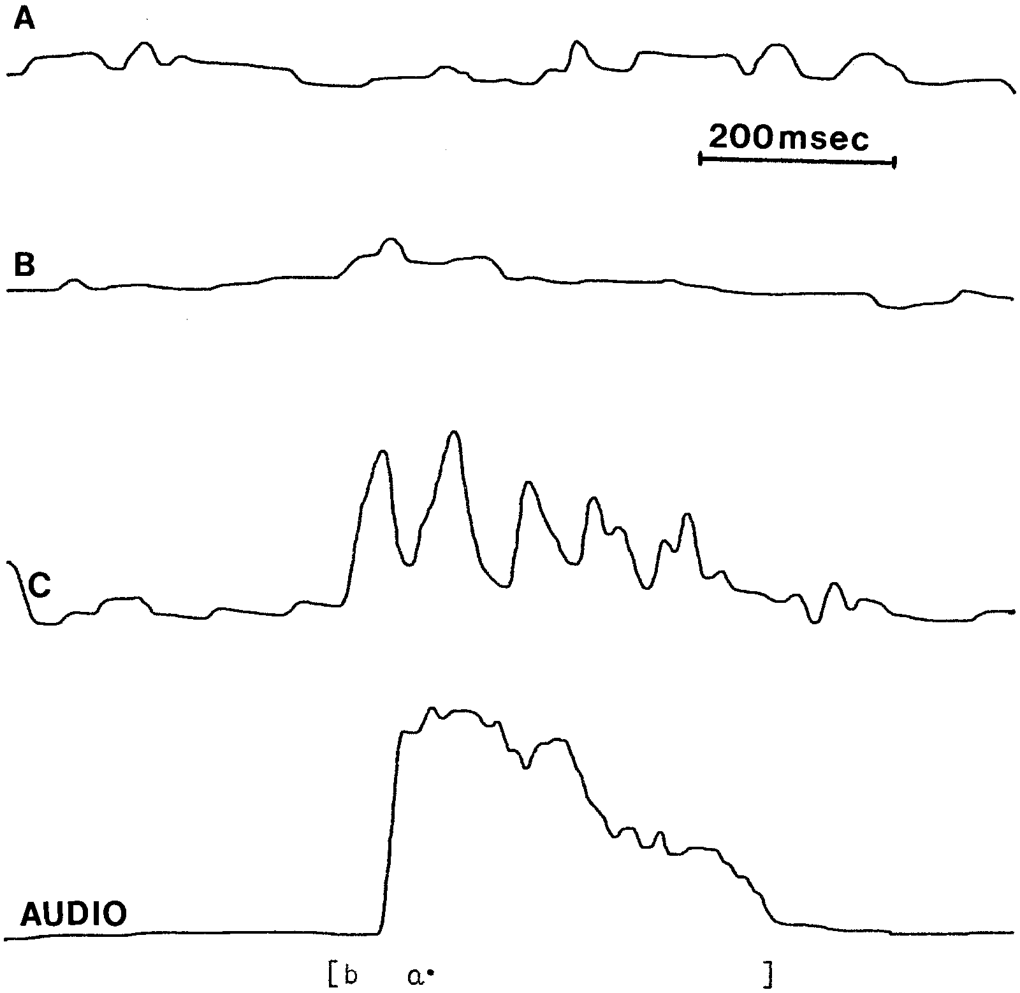


Figure 2.19a

Averaged emg signals from the electrode locations A, B, and C of experiment #2. CV syllable [bα·].

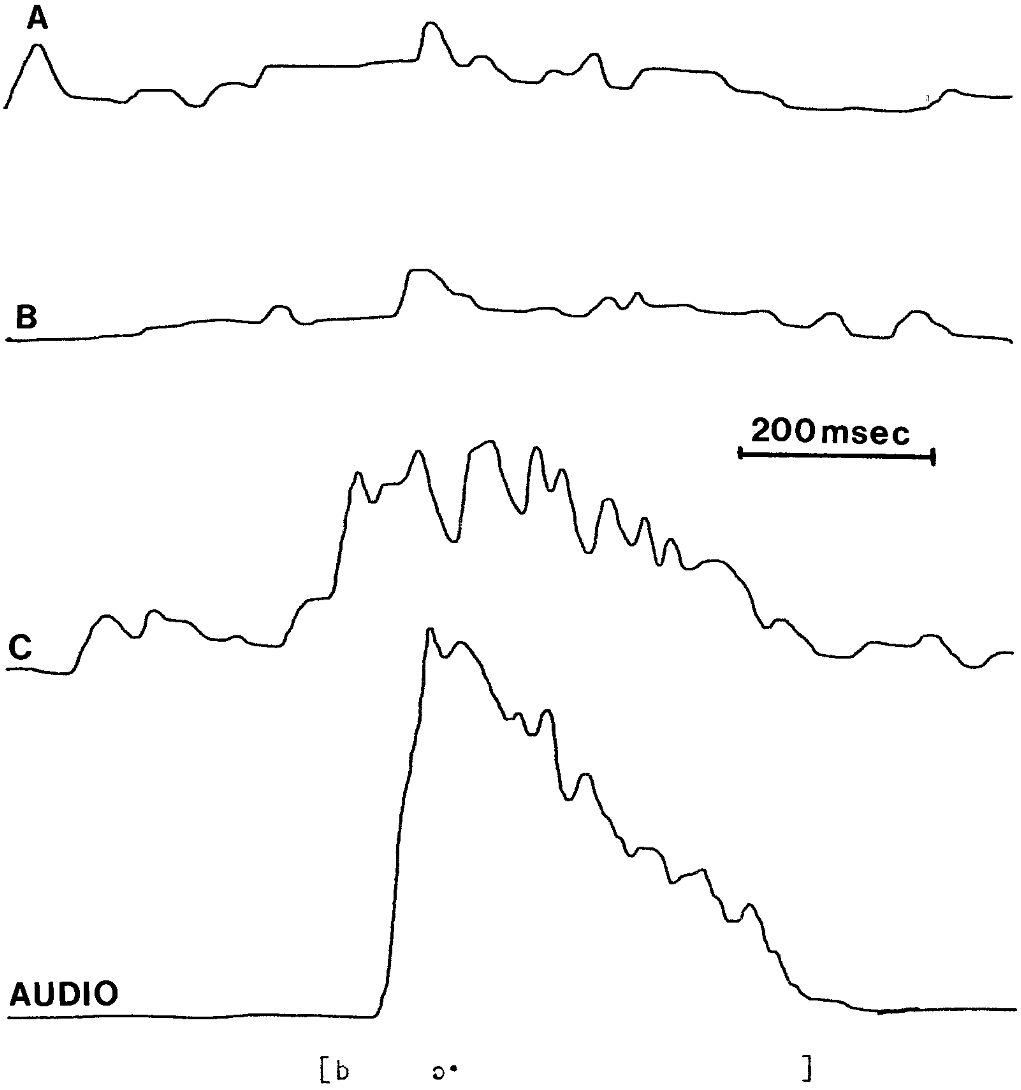


Figure 2.19b

Averaged emg signals from the electrode locations A, B, and C of experiment #2 for the CV syllable [bɔ·].

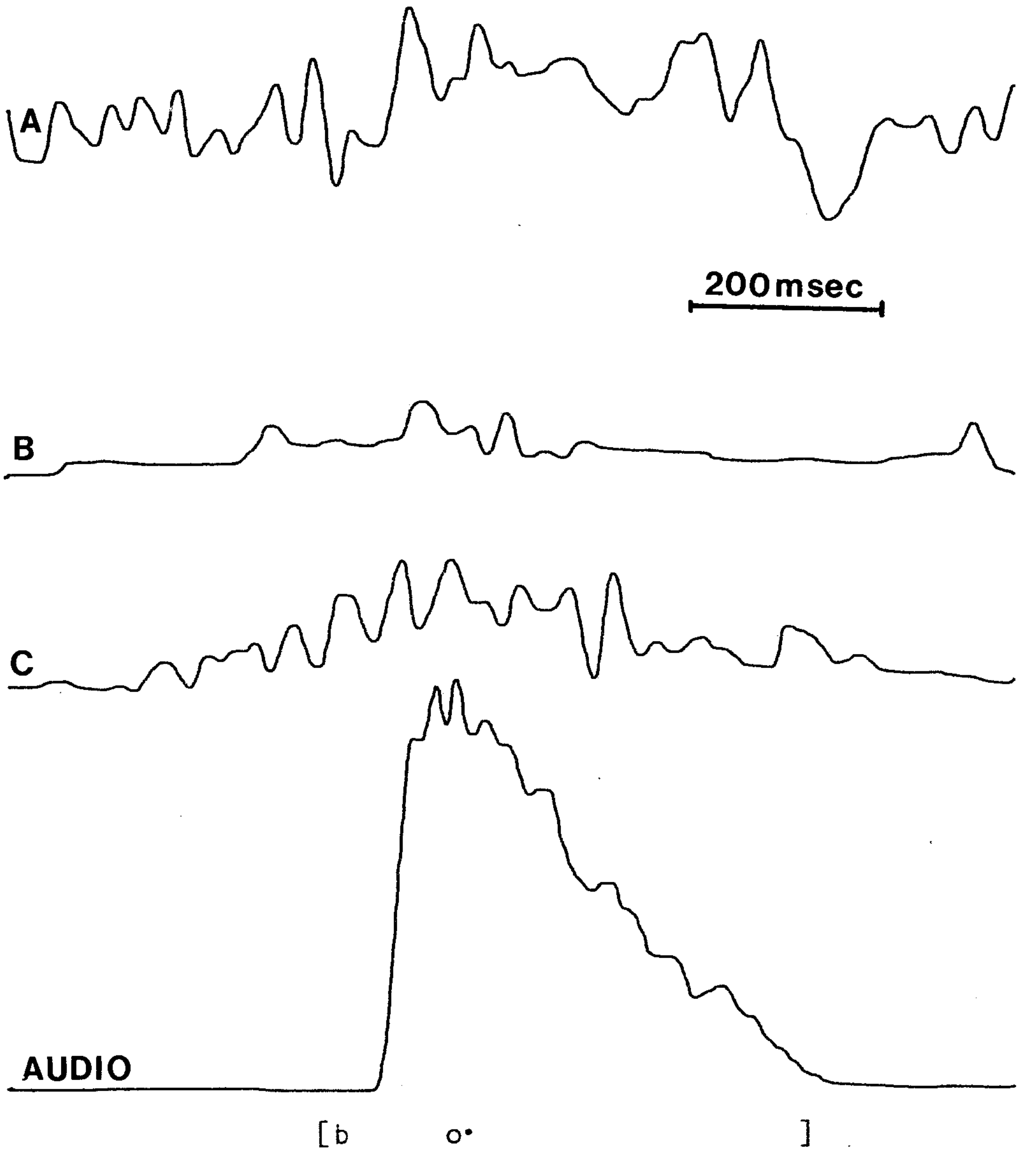


Figure 2.19c

Averaged emg signals from the electrode positions A, B, and C of experiment #2 for the CV syllable [bo:].

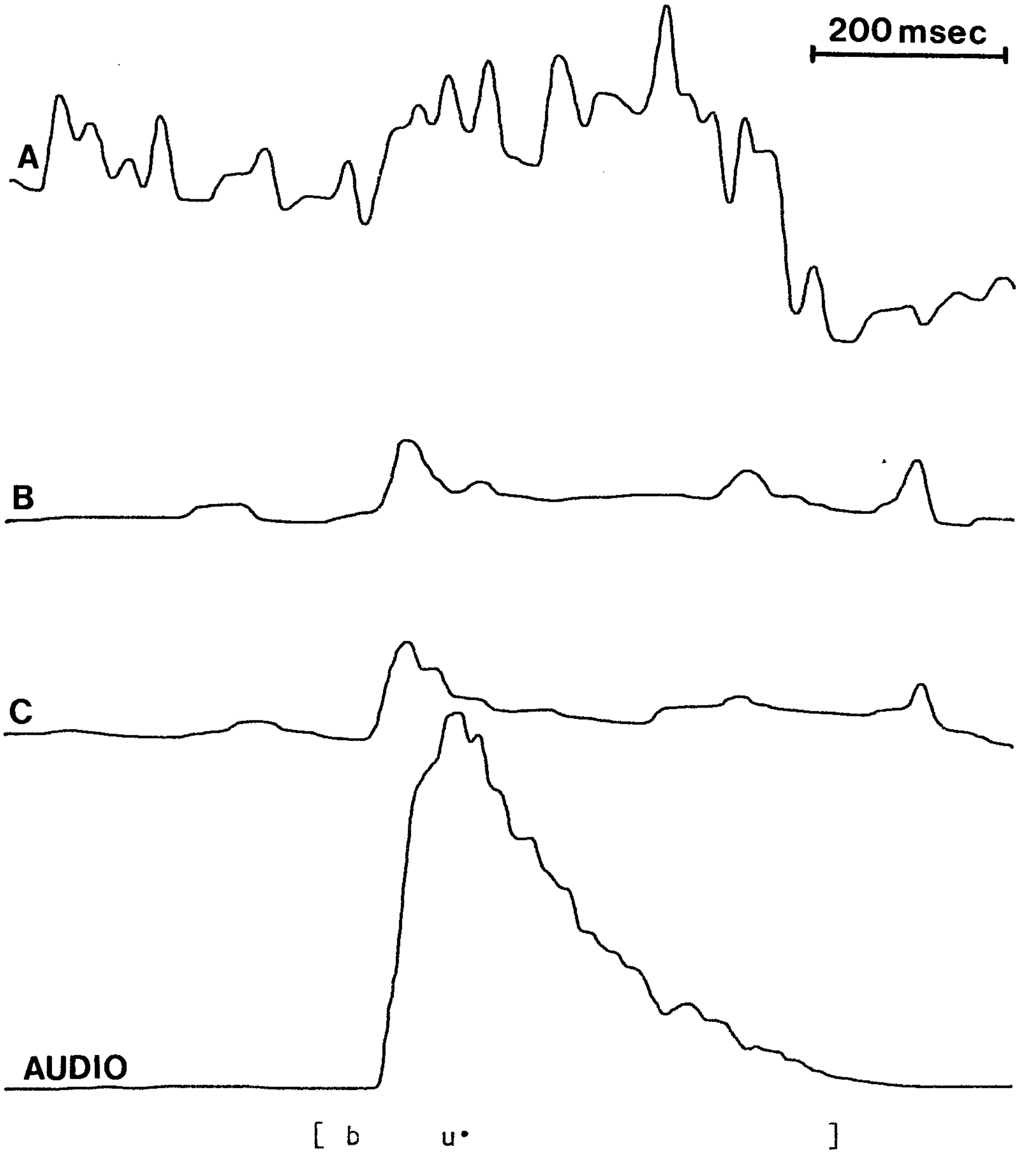


Figure 2.19d

Averaged emg signals from the electrode positions A, B, and C of experiment #2 for the CV syllable [bu:].

by careful reference to the raw data.

In the computer processing of the data in this experiment, the method of triggering the computer averaging procedure was changed slightly. As before, the audio signal was sensed by an advanced playback head on the tape recorder. The threshold level of the trigger channel was set so that it would respond to vowel peaks, but not to stop bursts or stop aspiration noise. When the first vowel of the frame ([ɩ]) was detected, the computer entered a delay loop. The length of the delay loop was adjusted so that it would end during the stop closure of the first consonant in the major word in the frame (just after the word "a"). The averaging window was then triggered by the next signal exceeding the trigger threshold, which was invariably the vowel of the major word in the frame. The averaging window was 1.54 sec in length. By decreasing the smear in the time domain between different tokens of one utterance, this procedure increases the reliability of the amplitude measurements of the averaged emg curves (Öhman 1966).

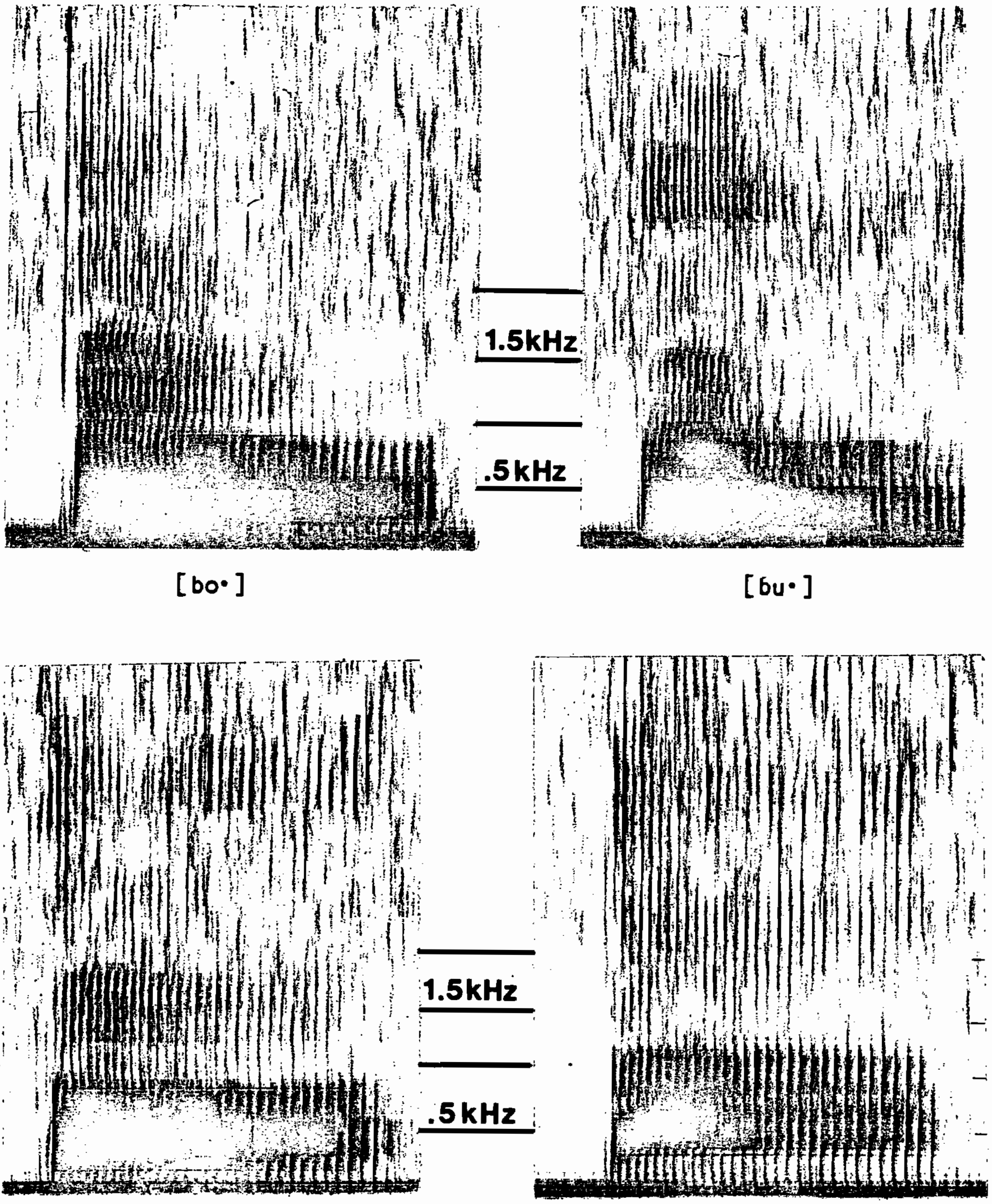
The averaged curves were plotted on a Hewlett-Packard Moseley X-Y plotter in the same manner as described above. Time calibration was also performed as described for the first experiment.

Results

The averaged and smoothed emg curves for locations A, B, and C for the vowels after the voiced labial stop /b/ are shown in Figures 2.18 and 2.19. The arrows on each curve indicate the onset and approximate offset of voicing for the vowel. X's indicate peaks caused by movement artefacts (usually near the point of release of the stop). A is most active in the high back vowel /u/. C is most active for the low vowels /a/, /ɑ/, and /ɔ/, the greatest activity being for /ɔ/, the vowel with the greatest degree of pharyngeal constriction. For /o/, all the channels show a small amount of activity. For /i/, there seems to be some initial activity on channel C, however this is clearly a movement artefact. Sound spectrograms of the back vowels /u/, /o/, /ɔ/, and /a/ are shown in Figure 2.20. Vowels with a high first formant (/a/ and /ɔ/) are those with the greatest degree of glossopharyngeal activity. The stability of the formant frequencies throughout the length of the vowels indicates the absence of diphthongization. The emg data are summarized in Table 2.4. In this matrix,

	i	a	ɑ	ɔ	o	u
A	1	1	0	0	1	2
B	1	1	0	0	1	0
C	1	2	2	2	1	0

Table 2.4



[bo·] [bu·]

[bɔ·] [bɑ·]

Figure 2.20

Sound spectrograms of some of the vowels produced by the subject PL of experiment #2. The horizontal lines indicate 500 Hz frequency calibration lines.

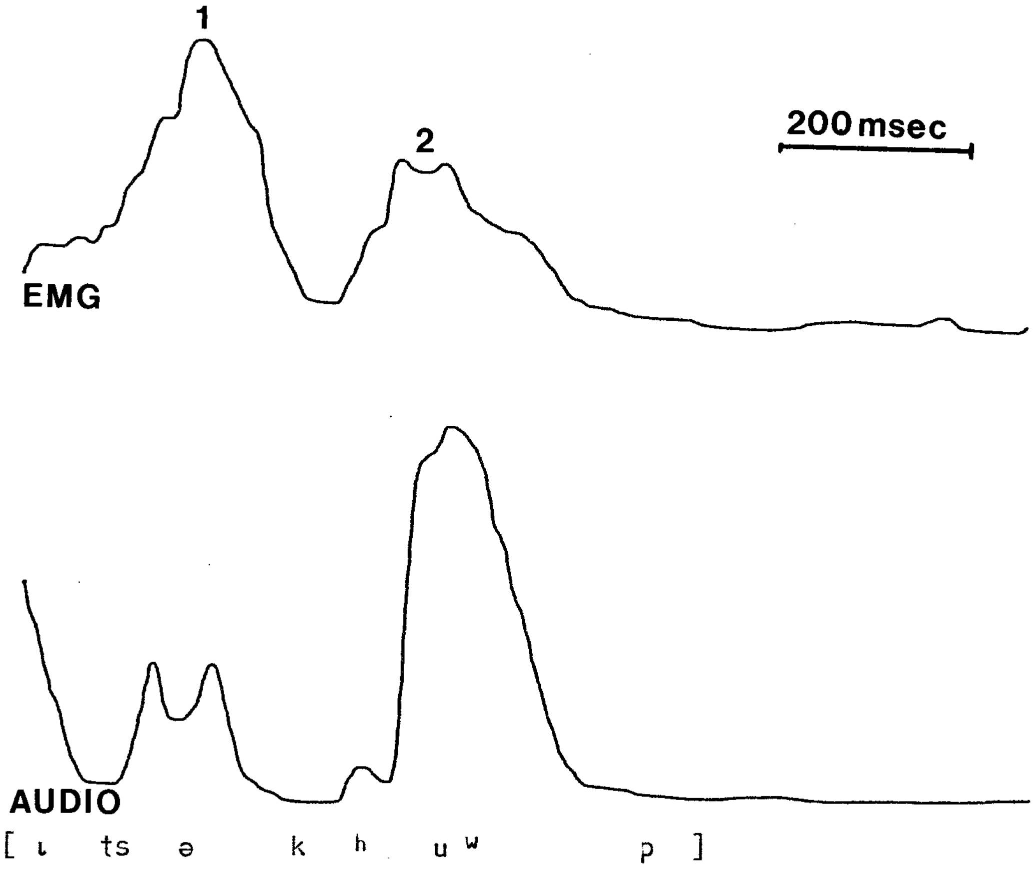


Figure 2.21

Averaged emg signals for the styloglossus muscle for the utterance "It's a *coop*." The first peak in the emg curve corresponds to muscle activity for the stop /k/, the second peak to activity for the vowel /u/ and the glide /w/.

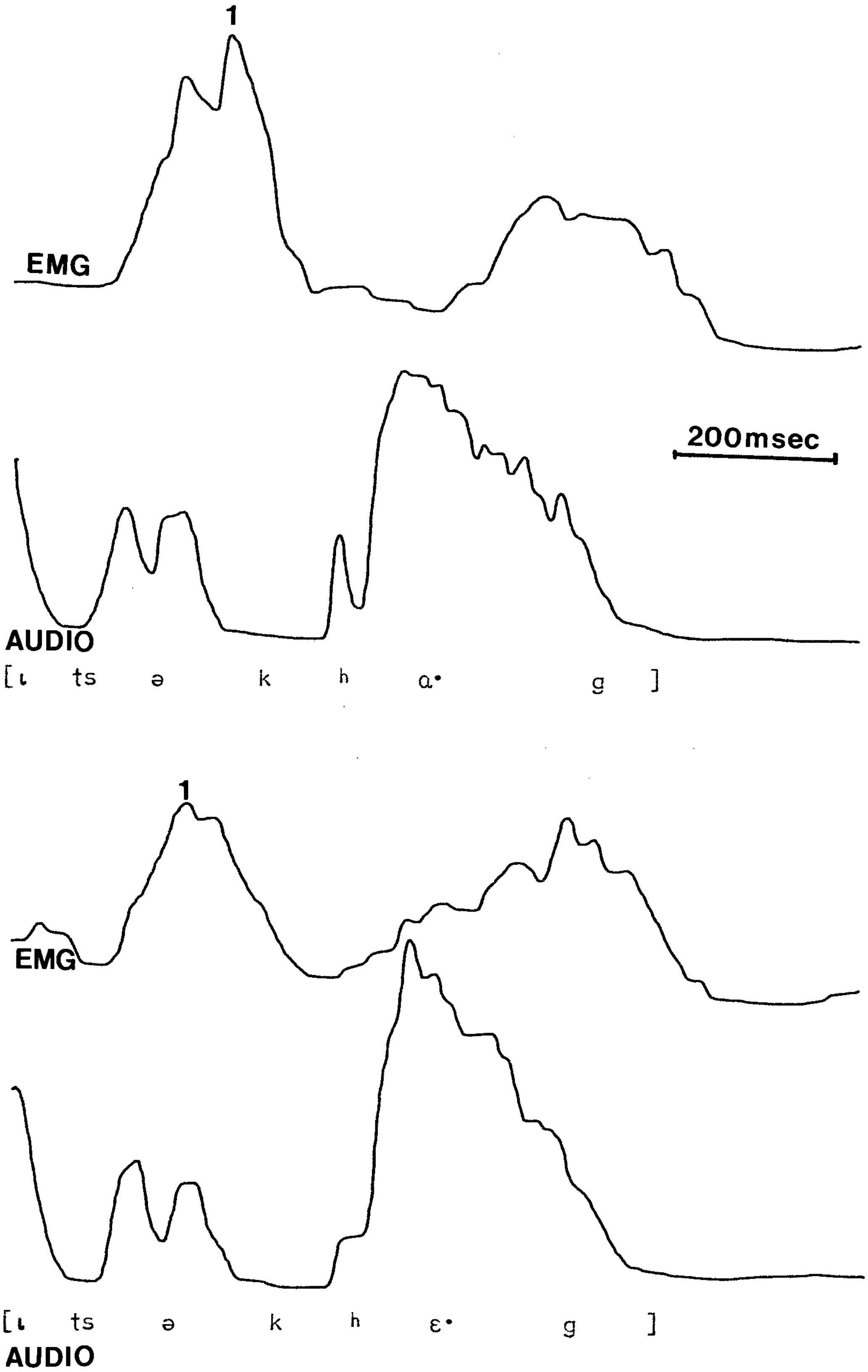


Figure 2.22

Averaged emg signals from the styloglossus muscle for the utterances "It's a <u>cog</u>" (top) and "It's a <u>keg</u>" (bottom). The initial peak in each curve is for the initial velar stop /k/.

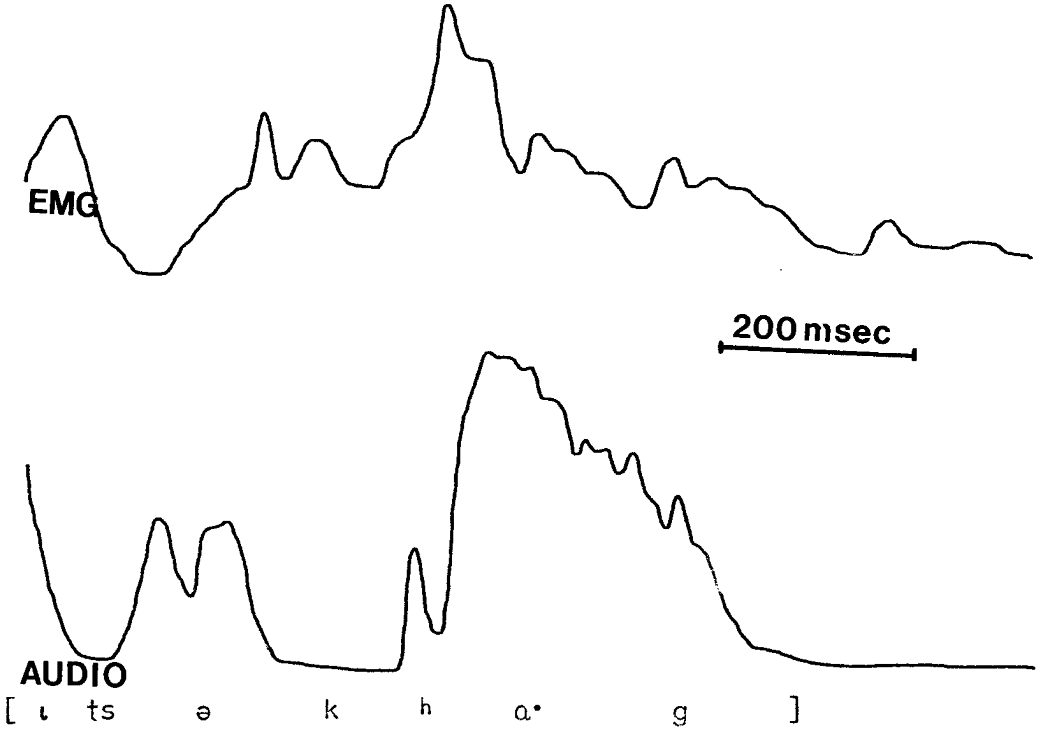


Figure 2.23

Averaged emg signal from the glossopharyngeal muscle for the utterance "It's a cog."

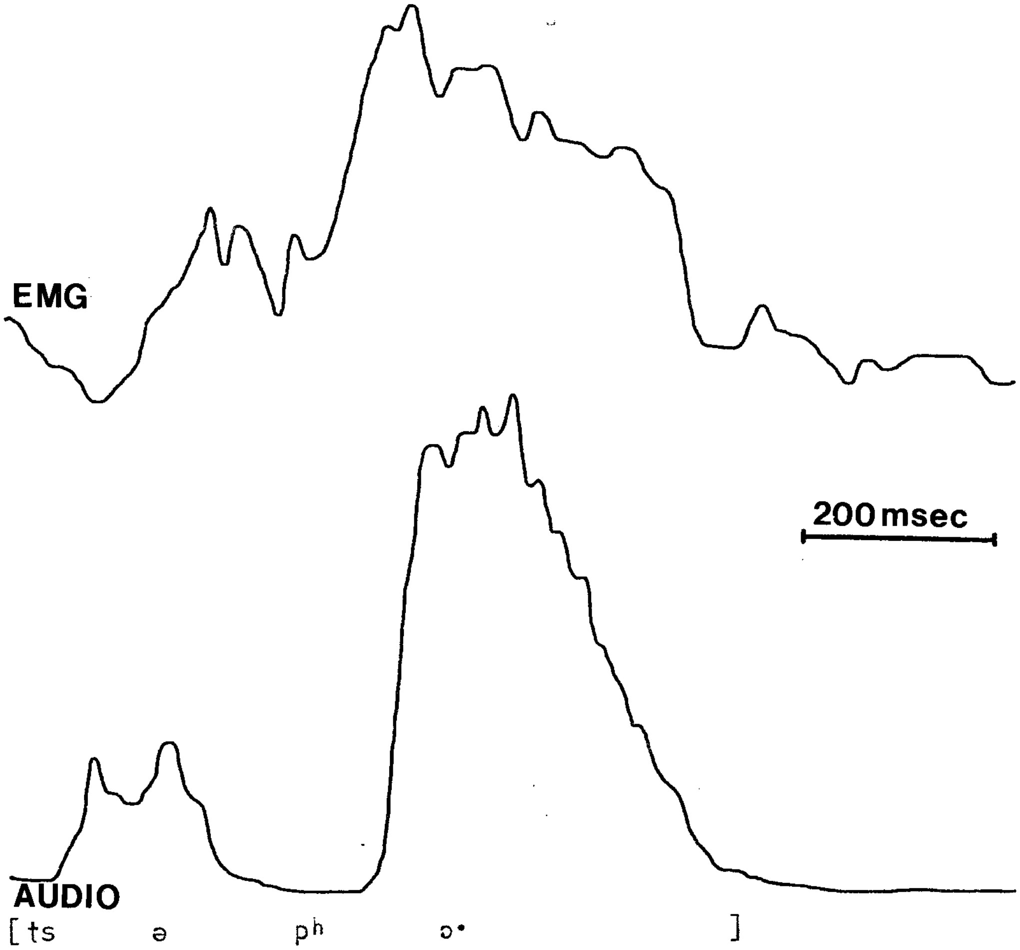


Figure 2.24

Averaged emg signal from the glossopharyngeal muscle for the utterance "It's a paw."

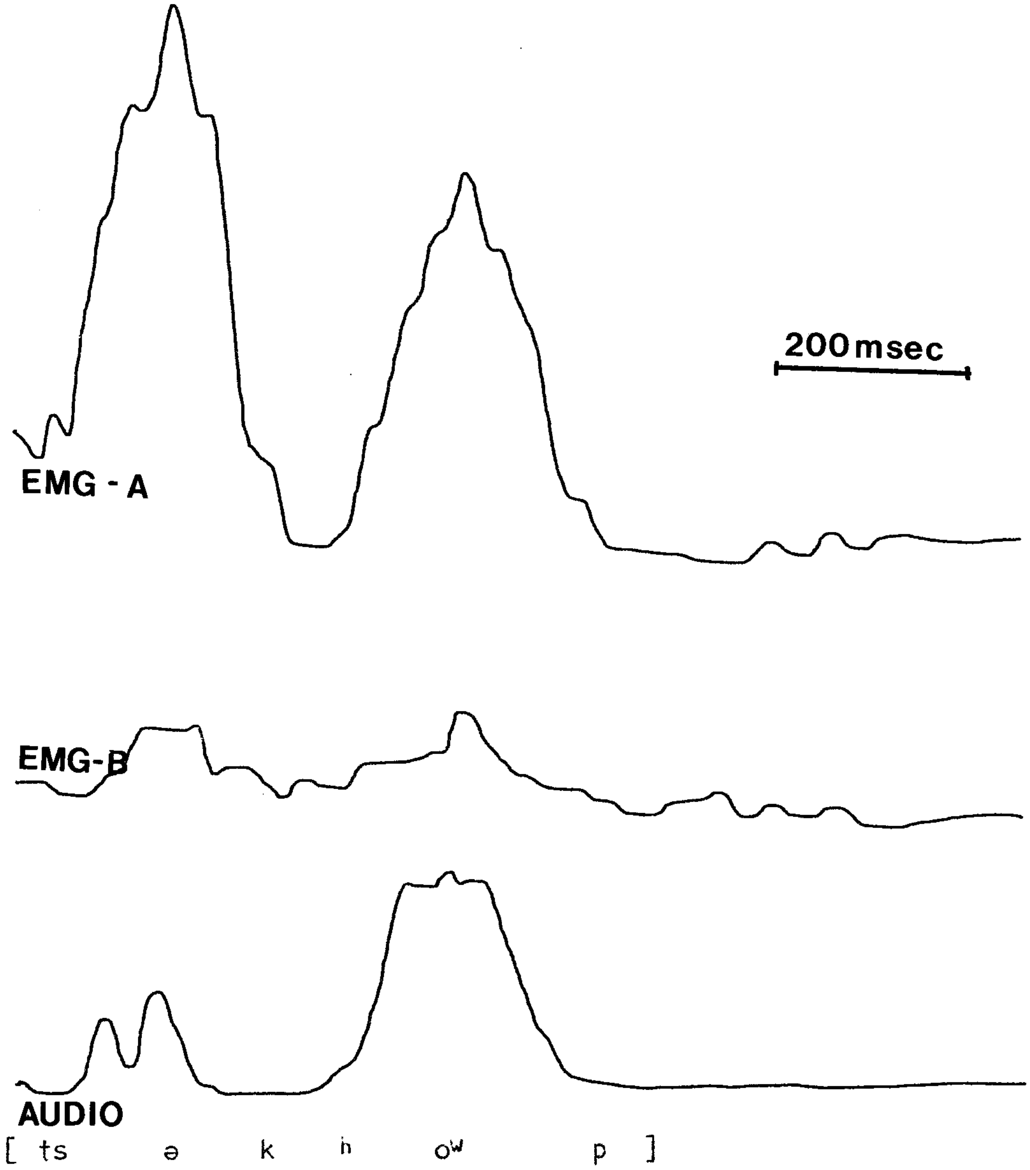


Figure 2.25

Averaged emg signals from the styloglossus muscle (A) and the glossopharyngeal muscle (B) for the utterance "It's a cope."

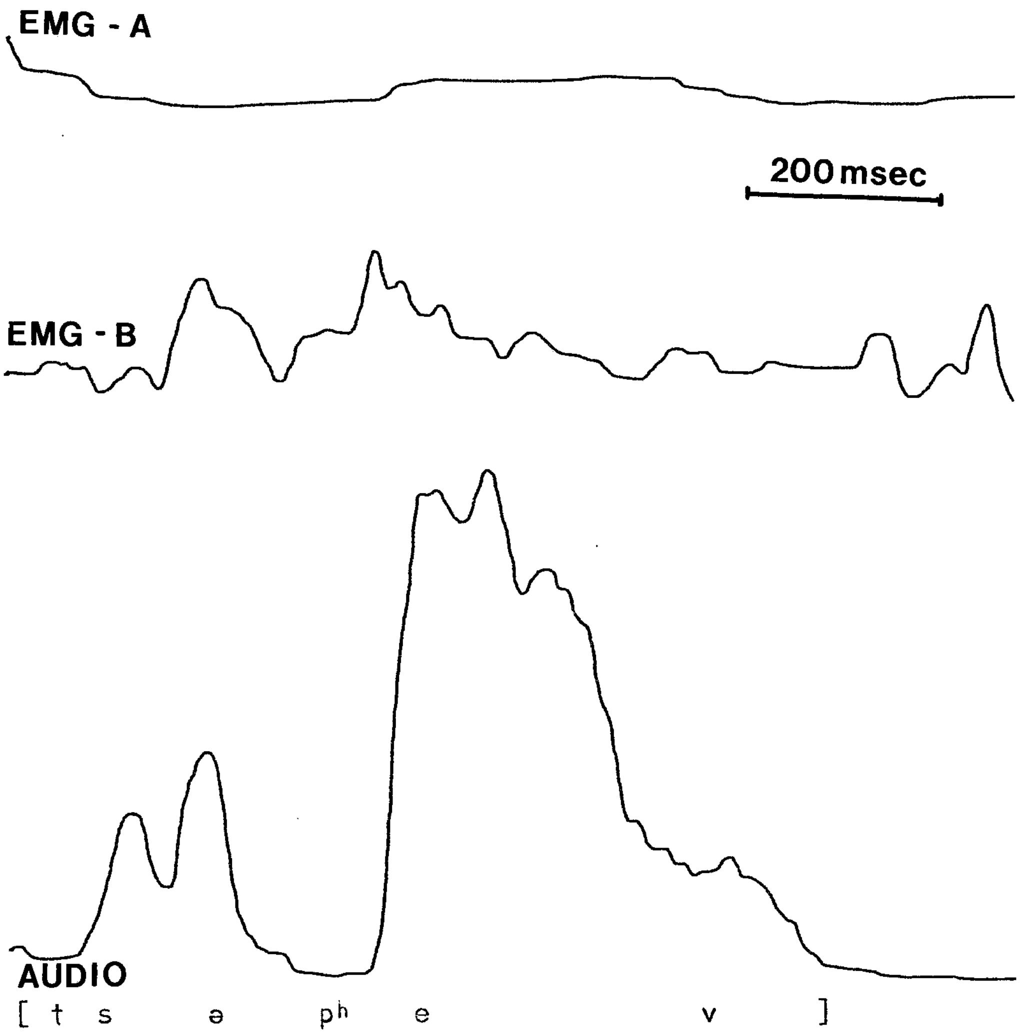


Figure 2.26

Averaged emg signals from the styloglossus muscle (A) and the glossopharyngeal muscle (B) for the utterance "It's a pave."

a 2 indicates a large amount of activity, a 1 an intermediate amount, and a 0 no activity.

It is clear that the glossopharyngeal muscle (location C) is active in all low vowels, while the styloglossus (location A) is strongly active only in the high back /u/. Analysis of the first part of the text supports these conclusions, and reveals some other interesting aspects of the behavior of these muscles. Figure 2.21 shows the averaged emg signal for the styloglossus in the utterance "It's a coop." This muscle is active for the velar closure, as indicated by the first peak, which leads the stop closure by 55 msec. The second peak is clearly associated with the vowel /u/. Note that the emg activity for this muscle is inhibited between the two segments. This inhibition is associated with the release of the velar closure. Since the tongue must be retracted quickly in the release of this segment, the styloglossus muscle is "turned off" momentarily, then resumes activity to stabilize the high back vowel position. This activity pattern is also noted in the posterior genioglossus muscle, in cases where there is a velar stop, followed by a vowel for which posterior genioglossus contraction is required. The posterior genioglossus is inhibited briefly, then resumes contraction. See Figure 2.11.

We would expect that the styloglossus should be more active in making a retracted velar closure than in making a fronted velar closure. It is. Compare the averaged emg curves for the styloglossus in Figure 2.22, where the items "keg" and "cog" are compared. The styloglossus activity for the velar stop in "cog" (/kɑg/) is considerably greater than in "keg" (/keg/).

Figures 2.23 and 2.24 illustrate the activity of the glossopharyngeal muscle in low vowels /ɑ/ and /ɔ/.

In the vowels in the environment after /b/, marked activity for the styloglossus was not noted in the production of /o/. The vowels in this series approached the quality of the cardinal vowels (which the subject is perfectly adept at producing), and were not diphthongized. See Figure 2.20. In Figure 2.25, the styloglossus activity for the /o/ in "cope" indicates a late, large peak. This peak is attributable to the diphthongization of the vowel in that environment. It can thus be established that the styloglossus is active in the labio-velar glide [w]. It does not contribute to the formation of the high front (or palatal) glide [j], as can be seen in Figure 2.26 of the item "pave", which *was* diphthongized. Activity of the styloglossus for the glide [w] is greater after a mid back vowel like /o/ than after /u/ (Figures 2.25 and 2.21).

The results of the second emg experiment can be summarized as follows:

1.) The emg signals from locations A and C represent, in all probability, the activity of the styloglossus and glossopharyngeus muscles, respectively.

2.) The styloglossus muscle is active in producing high back vowels and the velar part of the labio-velar glide [w].

3.) The styloglossus is also active in producing velar stops /k/ and /g/. The activity tends to be greater for a retracted velar (the stop before a back vowel) than for a fronted velar (before a front vowel).

4.) The glossopharyngeus muscle is active for consonants involving pharyngeal constriction, and for low vowels in general. This muscle appears to act as an antagonist to the genioglossus muscle.

5.) As with the muscles studied in the first experiment, definite peaks of activity and definite periods of inhibition are found. These peaks of activity and inhibitory periods can be uniquely associated with the segmental gestures of the utterance, or with certain features of these gestures. Large peaks are produced by the muscle at the onset of a gesture. Smaller peaks of the antagonist muscle are often, but not always, found at the "offset" of a gesture, even when the gesture for the following segment would normally provide sufficient return action.

GENERAL DISCUSSION OF EXPERIMENTS #1 AND #2

In this section we will summarize and discuss the activity of the muscles studied in these two experiments.

Genioglossus (anterior portion)

Of all the muscles studied, the anterior part of the genioglossus was the one which showed the least consistent patterns of activity. The activity for this muscle varied considerably from subject to subject. In fact there were clearly definable peaks which could be associated with speech segments in the activity of this muscle for only one subject of the four studied. In this subject, these peaks were largest for the alveolar stop consonants /t/ and /d/. It is unlikely, considering the location of this muscle, that it is primarily responsible for the articulation of these sounds, however. The actual closure of the vocal tract is probably effected by the superior longitudinal muscle, assisted by the mylohyoid. The more anterior fibers of the genioglossus probably serve only to bring the tip of the tongue forward in the oral cavity, positioning it under the alveolar ridge. Figure 2.10 shows that there is considerably greater activity of this muscle associated with the /t/ after a low vowel /ɑ/ than after the high vowels /i/ and /u/. This is to be expected, since the tongue must be moved farther to make the alveolar closure after a low vowel than after a high vowel.

Genioglossus (posterior portion)

The primary function of this muscle is to "bunch" the tongue, leading to the shape associated with front (/i/) and back (/u/) high vowels, and with the velar stops (/k/ and /g/). A certain amount of activity in this muscle is found during the mid vowels (/e/ and /o/). During the low vowels (/æ/, /ɑ/, /ɔ/, /ʌ/) the muscle is quiescent, however it becomes active towards the end of the vowel, where it serves to return the tongue body to the neutral vocal tract configuration for /ə/. This can be seen most clearly in the emg curves for the items /pɔp/, /pʌp/, and /pap/ in Figure 2.12b. These small peaks of activity must be associated with the reduced first vowel ([ə]) of the final word in the frame. It is clear, then, that the activity for a given muscle associated with a particular vocal tract configuration is a function of the previous state of the vocal tract. No posterior genioglossus activity is needed to reach the neutral ([ə]) configuration from an /i/ or an /u/ state, but activity is needed to reach this configuration from a lower (/ɑ/) or more backed (/ɔ/) configuration. Context dependencies of this sort are a feature of most recent models of speech production which generate a series of vocal tract configurations (Henke 1966; Öhman 1967a). Neurophysiological mechanisms which have been postulated to account for such effects will be discussed in detail in the next chapter.

The closest correlate of the posterior genioglossus activity in terms of the midsaggital representation of the vocal tract is the width of the middle part of the pharynx (line 3 in Figure 2.7). Measurements along this line for part of the utterances "it's a /pep/ again" and "It's a /pip/ again" are plotted in Figure 2.27. The dotted vertical lines represent the onset and offset of the vowel /e/.

Comparison of these curves with the emg curves for these CVC items in Figure 2.12a reveals the correlation between the posterior genioglossus activity and mid-pharyngeal width. Notice that the emg peak is later in the /e/ than in the /i/, corresponding to the later increase in pharyngeal width for the /e/. This activity of the genioglossus for /e/ is mostly associated, then, with the high front (palatal) off-glide on this vowel. Perkell (1969) found that the tense vowels in English involve greater displacement of the tongue body from the neutral position than their lax counterparts, and he hypothesized that this probably involves

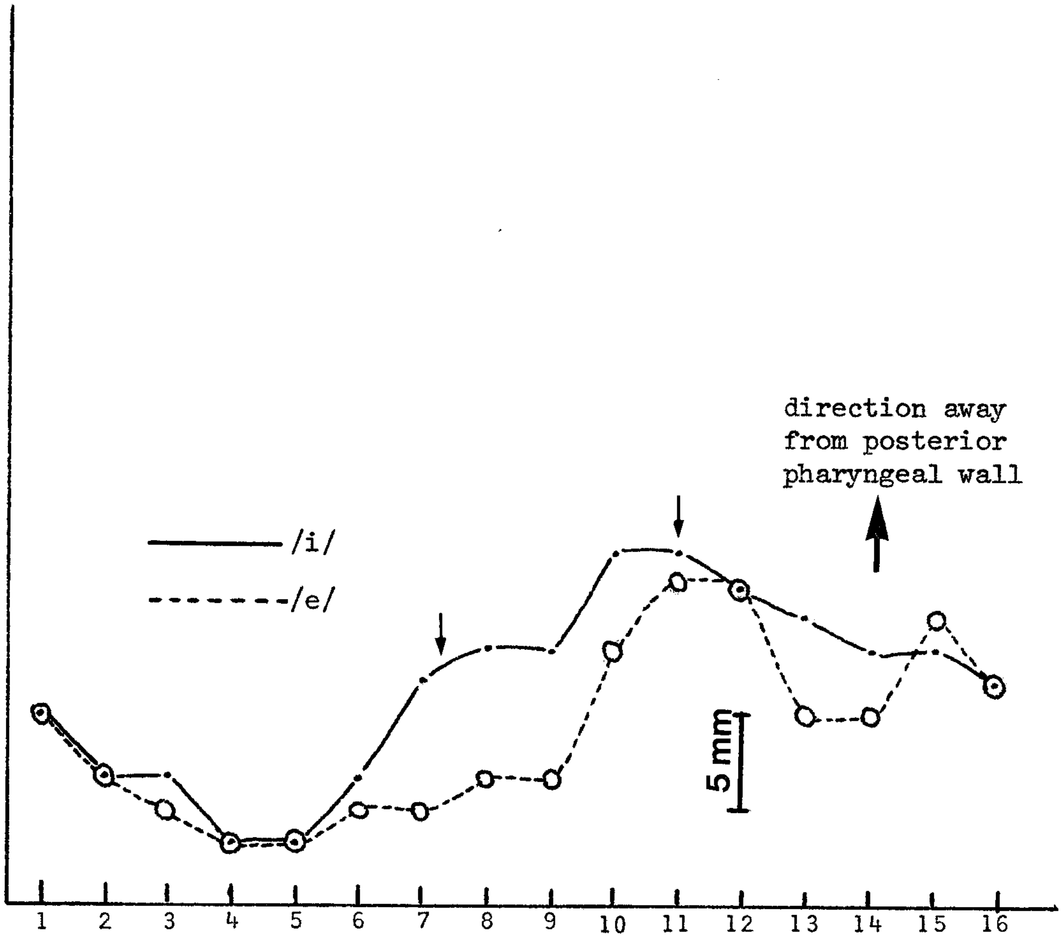


Figure 2.27

Width of the pharynx, measured along line 3 (see Figure 2.7) for the utterances "It's a /pip/..." (solid line) and "It's a /pep/..." (dotted line). The small arrows indicate the approximate points of onset and offset of voicing for the vowel /e/. The numbers on the abscissa are the frame numbers of the film (starting with the first frame after the stop closure for the /t/ of "It's").

greater activity of the extrinsic tongue muscles. Although this is probably true in general, the considerably greater activity of the posterior genioglossus for /i/, /e/, and /u/ (compared with /ι/, /ɛ/, and /ʊ/) found in this study must be primarily associated with the diphthongization of these vowels.

Perkell (1969) also found that the mid-pharyngeal width was greater during the production of voiced stops and fricatives than during their voiceless counterparts. This was attributed to a pressure regulation mechanism which increases pharyngeal cavity volume in order to minimize the build-up of pressure above the glottis. This increase in supraglottal cavity volume serves to maintain a trans-glottal pressure difference which aids in the voicing-through of the voiced obstruents. On the basis of Perkell's data, however, it could not be positively decided whether this increase should be attributed to passive expansion of the non-tensed pharyngeal cavity due to pressure build-up above the glottis, or whether the increase in pharyngeal width was due to an active, muscular mechanism.

This widening of the pharynx for voiced obstruents has been verified by Kent and Moll (1969), and was also observed in this study. In Figure 2.28 the mid-pharyngeal width along line 2 (see Figure 2.7) is shown for the utterances "It's a /tat/..." and "It's a /dad/..." In addition, the vertical displacement of the hyoid bone is also shown. The approximate points of closure for the stops in the words /tat/ and /dad/ are indicated by the arrows. Just after the closure for the initial stops, the mid-pharyngeal width is approximately 11 mm greater for /d/ than for /t/. This effect is not found during or just after the closure for the final stops in these words.

Kent and Moll (1969) claim that the increase in pharynx width in voiced obstruents is attributable to muscular activity, rather than to a passive reaction to the increase in oral pressure during the lax voiced obstruents. The most likely means by which mid-pharyngeal width could be increased would be contraction of the posterior portion of the genioglossus muscle. As noted above, contraction of this muscle is correlated with increased pharynx width in the high, tense, diphthongized vowels /i/ and /u/. However, in this study, it was found that the activity of the posterior genioglossus muscle during the voiced alveolar stops was not significantly different from the activity during the voiceless stops. In fact, the voiceless stops often showed slightly greater activity (see Figure 2.14). This was true for all 4 subjects studied. Kent and Moll, however, hypothesize that the increase in mid-pharyngeal width is due to the depression of the hyoid bone. Figure 2.28 shows that the hyoid bone in the present study was about 4 mm *higher* just after the initial /d/ of /dad/ than after the initial /t/ of /tat/. This was true of the final stops also.

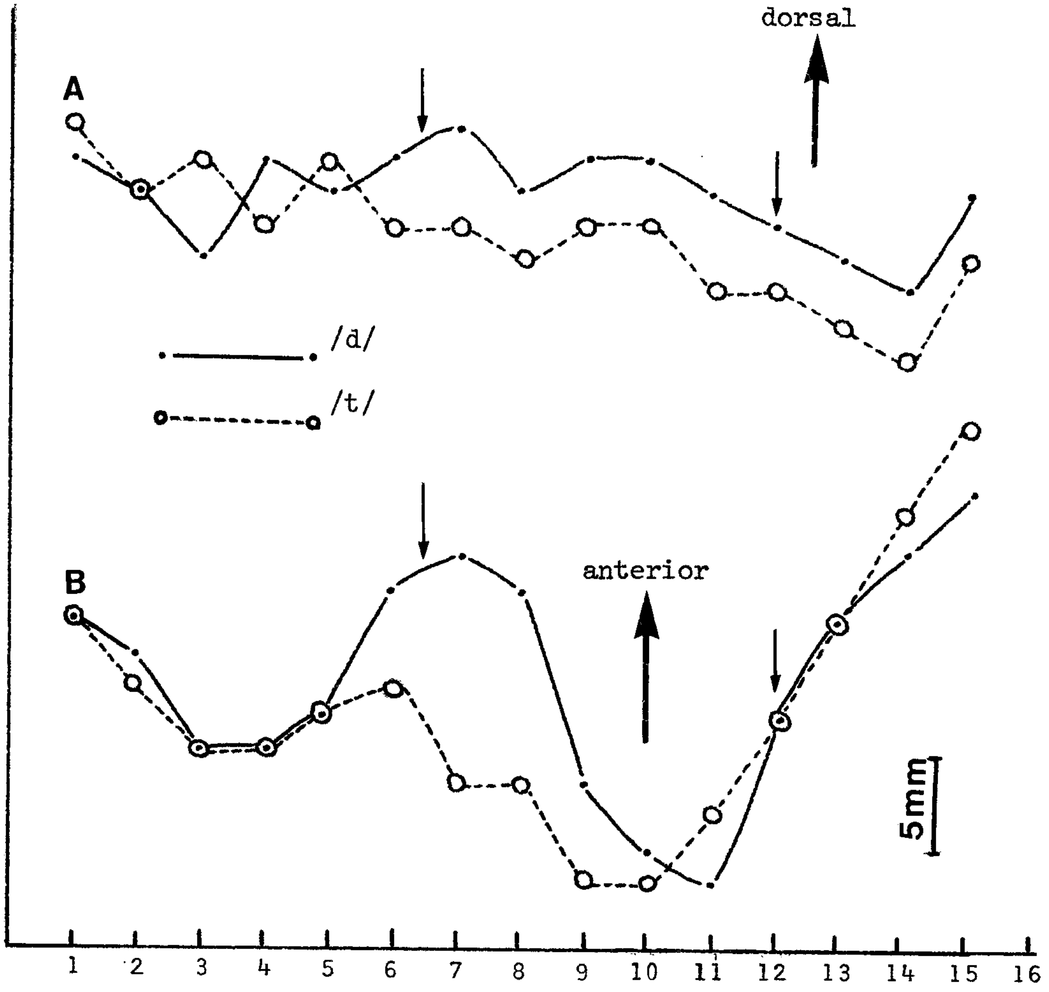


Figure 2.28

Elevation of the hyoid bone (A) and mid pharyngeal width (B) for the items "It's a /tat/..." and "It's a /dad/..." The small arrows represent the approximate center of the closure for the stops in the /CVC/ words. The numbers on the abscissa are the frame numbers of the film (starting with the first frame after stop closure for the /t/ of "It's"). The mid pharyngeal width was measured along line 2 (see Figure 2.7).

On the basis of the data reported here, we conclude that the widening of the pharyngeal cavity during voiced stops is primarily due to the passive reaction of the non-tensed vocal tract. The fact that the difference in mid-pharyngeal width is not found in the final stops of the CVC words is probably due to two factors: (1) the shorter closure duration of the final stops, and (2) the anticipatory widening of the pharynx for the velar stop of the last word in the frame (a*g*ain).

Mylohyoid

The mylohyoid muscle, as indicated above, is active mainly for the velar stops. It also shows some activity for the alveolar stops, and a very slight degree of activity for high front vowels.

Styloglossus

The styloglossus muscle serves to raise and retract the tongue during velar stops and high back vowels and glides. It is not active during the production of low back vowels, indicating that it does not act to simply retract the body of the tongue during vowel production, but only to retract and lift the tongue. For the velar stops, there is greater styloglossus activity for the retracted allophone of /k/ (the one before /ɑ/) than for the "advanced" allophone (before /ɛ/). In the production of these velar stops, the styloglossus acts as one member of a synergistic complex of muscles. The posterior genioglossus and the mylohyoid (along with the palatoglossus) comprise the other members of the complex. It seems quite likely that the activity of these muscles alone could serve to account for the shape of the tongue observed during velar occlusion. In this sense, they could be thought of as constituting a "dorsal articulatory complex" for consonants, similar to the "dorsal articulator" proposed by Öhman (1967a).

It is not possible, for reasons stated above, to record emg signals from isolated intrinsic muscles of the tongue. It seems likely, however, that the contribution of the intrinsic muscles to the production of dorsal articulations is rather minimal. Perkell (1969) has hypothesized that the extrinsic muscles serve only to position the body of the tongue during consonant articulation,while the actual closure is made by the intrinsic muscles. This would appear to be the case for dental and pre-palatal consonants, as noted above. However it is clearly not the case in the production of the velars. The activity of the extrinsic muscles studied here is important in the formation of velar consonants. The electromyographic data thus provide considerable support for the claim that different articulatory complexes are involved in the production of pre-palatal and post-palatal consonants. The distinction of apical and dorsal articulators made by Öhman (1967a), and the feature *coronal*, proposed by Chomsky and Halle (1968), appear to have a real physiological basis. These matters will be discussed in greater detail in Chapter 4.

Glossopharyngeus

The glossopharyngeal part of the superior pharyngeal constrictor muscle serves to retract the body of the tongue in general, and to form a pharyngeal constriction in the case of pharyngeal fricatives. This muscle, along with the medial portion of the genioglossus, is also probably active in forming the vocal tract shapes for pharyngealized consonants.

Chapter 3

The results cited in the preceding chapter give a general picture of the extrinsic muscular control of the tongue in speech. In describing these results, emphasis was placed mainly on the role of individual muscles in producing the tongue configurations found in a limited set of segment-types. In this chapter, the results of the experiments described in Chapter 2 will be discussed in more detail, and will provide the basis for a general discussion of some aspects of speech production models which have been proposed in the recent literature.

One of the major sources of difficulty for those interested in speech production lies in the nature of recent neuro-physiological research on motor function. Excellent work has been done on cerebellar function (Eccles 1967) and on the relation between motor cortex activity and motor behavior (Evarts 1967), and we know much about the function of lower-level reflex control mechanisms (Matthews 1964). Almost all of this work, however, has been concerned with the spinal muscles of animal preparations. There are obvious reasons for this. No sane person would conduct this necessarily traumatic research on human subjects. Nevertheless, one finds oneself wishing that more physiologists would leave intact their cats; let these poor animals live to hunt another day; and turn to at least speculation concerning the infinitely more interesting aspects of motor performance found in human language.

Another, more important, source of difficulty in proposing and evaluating models of speech production may lie in a tendency to postulate either/or models: "Either the articulatory gestures of speech reflect a target-directed displacement component, or they are framed in terms of amount of contraction of a set of muscles." As Evarts (1967, p. 245) has said:

> It is often the case that the nervous system solves problems by both of two alternative approaches and does not rely on either approach exclusively, and it is dangerous to state questions that require the acceptance of one mode of functional organization to the total exclusion of the other.

Many of the questions that have been proposed in attempts to construct models to account for the motor organization of speech tend to be framed in terms of this sort of all-or-nothing attitude. The questions are familiar. Is speech behavior open-loop or closed-loop? Do the actual motor commands to the muscles used in speaking remain invariant for the same phoneme in different contexts? If they are not invariant, is this a result of some type of central programming or of peripheral reorganization? Are the muscles used in speaking controlled in terms of central commands which dictate a particular spatial configuration of the vocal tract, or are they controlled by commands dictating a particular amount of muscular contraction? These are all exceedingly complex issues, and most of them cannot be posed as either/or questions. Consider as an example the matter of feedback. It is transparently obvious that feedback of some sort is used in speech. If it were not, complete deafferentation and total hearing loss would result in no deterioration, either long-term or short-term, in speech behavior. However even partial deafferentation can cause some deterioration in motor abilities in speech (Ringel and Steer 1963, Ladefoged 1967a) and total hearing loss results in eventual alteration of speech patterns. It is probable that there are no motor processes in mammals which are devoid of sensory feedback of some sort, or at least the possibility of feedback. In discussing the motor control of speech, it is necessary to determine what sort of feedback is or can be used, and at what levels of the process it is adopted.

In the ensuing discussion, we shall examine closely a number of questions similar to those posed above. It is important to note that these questions are not independent of each other, and cannot be fully discussed and resolved (if they can be resolved at all) independently.

THE NEURO-MUSCULAR INVARIANCE HYPOTHESIS

Since language can be described in terms of phonemes, or in terms of features of phoneme-sized units, it is an apparently logical step to infer that these units are stored somewhere in the brain in terms of the motor patterns that underlie the articulatory behavior of the vocal tract in speech. Stated in an over-simplified way, the process of speech production could then be viewed as the concatenation of these units by semantic, syntactic, and phonological performance rules, the assignment of a temporal sequence of the units (the rhythm and rate of the utterance), and the transmission of these motor patterns to the speech muscles via the cortical motor neurons and the cortico-bulbar and cortico-spinal motor pathways. A corollary of this view, which can be called the "neuro-muscular invariance hypothesis," might be that the motor unit activity of the muscles principally involved in the production of a particular phoneme should remain essentially invariant in the production of that phoneme, independent of the immediate phonetic context. Since it is obvious that the articulatory gestures of speech do vary depending upon phonetic context, adherents of the neuro-muscular invariance hypothesis are forced to explain this variation in terms of the mechanical properties of the articulators. Thus it would be hypothesized that although the neuronal commands to the muscles are essentially the same in most of the muscles involved in the production of a phoneme, the inertia and damped response of the articulators themselves could produce different observed gestures depending upon the different positions of the articulators in the immediate phonetic context. This view of the production of speech has been tentatively proposed by some members of the Haskins Laboratories group (Cooper 1966, Liberman et al. 1967). In fact, it is not unreasonable to claim that this was, until very recently, the most common view of the articulatory behavior of the vocal tract. For example, Stevens and Halle (1967, p. 90) write:

> When [the instructions for particular types of behavior of the speech-generating mechanism] are executed, the interaction between different physiological structures, each possessing its own sluggish response characteristics, will naturally produce a quasi-continuous gesture in which the discrete instructions initiating the gesture are no longer always discernible as distinct components.

The hypothesis that the neuro-muscular behavior of the vocal tract is in terms of phoneme-invariant commands to the muscles is a difficult one to evaluate, however. One problem, mentioned by Fromkin (1968), is presented by the abstract nature of current phonological theory. In generative phonological theory a phoneme is a highly abstract unit, which does not necessarily bear a close relationship to some directly corresponding segment in the pronunciation of the utterance (Chomsky 1964, Postal 1968). In fact, some varieties of generative phonological theory (Chomsky and Halle 1968, Hyman 1970) allow the setting up of phonemes in a language which are never realized in the pronunciation of an utterance in that language. Clearly, some much less abstract level of linguistic representation offers more fruitful possibilities for the evaluation of the motor equivalence of segment types. In this discussion, we will consider that the neuro-muscular invariance hypothesis should apply to what may be called *intrinsic* allophones. These are defined, intuitively, as allophones of a phoneme which are closely related phonetically, such as the velar stops of the words *key* [k͔] and *car* [k]. Extrinsic allophones, on the other hand, are realizations of a phoneme which are articulatorily somewhat more distinct, such as the clear and dark allophones of /l/ in English: *leaf* [lijf] and *feel* [fijɫ] respectively. As Ladefoged (1967b) pointed out, this distinction can also be captured more formally in terms of the specifications necessary to control an articulatory model of the vocal tract (such as the one proposed by Henke 1966). If the articulatory distinctions between two or more realizations of a phoneme can be accounted for by a single target specification of the vocal tract and a coarticulation function of the sort described by Öhman (1967a) then the realizations are considered to intrinsic allophones. If different target specifications are called for the realizations are extrinsic allophones. While retaining these definitions Ladefoged (1970) now considers it preferable to change the terminology and refer to coarticulated (formerly intrinsic) and target (formerly extrinsic) allophones. These two classes of allophones will be considered in further detail in Chapter 4.

The only means available at present for evaluating the neuromuscular invariance hypothesis is recording the electromyographic signals representing the motor unit activity of the muscles involved in producing a particular segment. The emg studies which have been done to date demonstrate rather convincingly that there is a considerable degree of variation in the emg signals from muscles principally involved in the articulation of intrinsic allophones of a phoneme. Fromkin (1966), Ohala and Hirano (1967), Öhman (1967b), MacNeilage and DeClerk (1969), and Fritzell (1969) have all reported these effects.

In addition, the emg data in this study do not support this hypothesis. For all the muscles studied, we found context effects for similar segments which were apparently determined by the phonological and phonetic identity of both the preceding and following segments. The best example in our data is found in the emg signals for the segment /k/ in different environments for the mylohyoid, genioglossus, and styloglossus muscles. The behavior of the mylohyoid muscle can be taken as an example. It was found, in all of the subjects studied, that the averaged emg signals for the /k/ gesture in this muscle differed for the advanced and retracted allophones of /k/ . There was greater activity for the [k] before [ɑ] and [u] than for the [k˂] before [i]. Similar differences were found for the final /k/ in the CVC syllables investigated.

One of the problems associated with performing a viable electromyographic test of the neuro-muscular invariance hypothesis is the determination of the degree of variation possible before the gesture can be deemed to be different. There are no standard procedures for determining the amplitude of interference pattern emg signals. An inspection of averaged curves such as those presented in Chapter 2 above can give some indication of the degree of difference, however the differences cannot always be evaluated in a straightforward manner. There are a number of linguistic factors which can interact to produce widely different neuromuscular events for the same phoneme. For example, Harris et al. (1969) have shown that the muscular action involved in the production of labial gestures differs depending upon the prominence of the syllable. For some speakers, there may be an increase in lip muscle activity in stressed syllables. Furthermore, they have shown that different speakers may use their muscles differently to produce phonologically and articulatorily similar gestures. Similar effects were noted by Fritzell (1969) and in this study.

In order to gain more information concerning the effects which can influence the amplitude of *averaged* emg signals for similar articulatory gestures, we will take a closer look at some of the data of Experiment #1 reported in Chapter 2. For the most part, we will be concerned with factors which can affect the signals produced by a single subject during the same experiment. First, however, we will discuss some factors which could influence the emg signals produced by different subjects, or by the same subject on different occasions, when one attempts to examine the electromyographic activity of a single muscle.

Cross subject comparisons

In order to compare the rank order of emg signal amplitudes across subjects, it is necessary to assume that (1) the emg activity from the same muscle is being sampled in each subject, and (2) the emg activity was recorded from motor units of the same type, or that a sufficient number of motor units was sampled to guarantee comparable records. Condition (1) was, we feel, met in the first experiment described in Chapter 2, involving recording from the genioglossus and mylohyoid muscles. Condition (2) requires some further explanation.

Most skeletal muscle contains two different types of muscle fiber (Henneman and Olson 1965). Some motor units are made up mostly of small diameter, red, "slow" fibers. These fibers are adapted to firing at relatively low frequencies (1/sec to 10/sec). Other motor units are composed mainly of larger, pale, "fast" fibers, which are better adapted to firing at higher frequencies (10/sec to 100/sec). The fast fibers reach their maximum contractile force more quickly, and are able to exert a higher degree of effort than the slow fibers. On the other hand, they tend to fatigue more quickly. Most muscle is made up of motor units containing both types of fiber, however one type is usually predominant. This allows muscles to be classified along a scale of "fast" to "slow". Typical of the fast muscles in the human body are the extrinsic eye muscles. The soleus, a leg extensor, is a slow muscle. Tokizane and Shimazu (1964) have shown, using functional criteria, that the genioglossus muscle in man is predominantly made up of fast fibers. They recorded single motor unit activity at different electrode locations under different conditions of force of contraction. They studied distributions of the averaged interspike interval plotted against the variance of the intervals, on the hypothesis that slow fiber motor units (tonic neuro-muscular units, in their terminology) should show more variance at higher firing rates, while fast-fiber motor units (kinetic neuro-muscular units) would have a higher variance at low frequency firing rates. They found that the distribution of units in the same muscle was bimodal in most cases when examined in this manner. Some muscles, such as the orbicularis oculi and the genioglossus contained mainly kinetic units, while typical "slow" muscles such as the leg and thigh muscles had more tonic units. Of all the muscles that they studied, the genioglossus was among the "fastest," just slower than the orbicularis oculi, an extrinsic eye muscle. This is of interest in view of the claim made by Perkell (1969) that the extrinsic tongue muscles such as the genioglossus are "slow" muscles.

The raw emg data in this study which produced the large, sharp, well-defined peaks in the averaged curves were always interference pattern signals composed of 5 or more motor units. In the cases where lesser amounts of activity were found, however, it was often the firing of only one motor unit that produced the averaged emg curves. An interesting example of this is found in the behavior of the mylohyoid muscle during vowels. Figure 3.1 demonstrates a single token of emg in the mylohyoid for the utterance "It's a /kik/ ...". The interference patterns in the peaks for the /k/'s can be seen. During the vowel /i/ , only two motor units appear to be firing, at a rate of about 20/sec. Thus it is possible that the small amount of activity of the mylohyoid noted during some vowels is a result of the firing of "slow" or tonic motor units, while the consonant activity recruits the faster, phasic units in this muscle.

In general, it is unlikely that this distinction between slow and fast units would contribute to significant differences in recorded activity for the same gesture from subject to subject, since the fibers of these two types of motor units are not grouped together, but are randomly scattered within most muscles.

Factors affecting the amplitude of peaks in averaged signals

It is a well-established fact that single tokens of physiological signals such as emg cannot be compared meaningfully (Fromkin and Ladefoged 1966). It is necessary to average at least 20 tokens of a particular gesture before meaningful comparisons can be made. Nevertheless, it is still the case that gestures involved in speaking will be influenced by factors which have nothing to do with the immediate phonetic context. In order to examine these factors in detail, we will look at the emg activity of Experiment #1 of this study for the items in the frame as well as the items in the various CVC syllables.

In this aspect of the study, the emg data from one subject only was used. This subject (TS) was chosen for several reasons: (1) he was the subject for the X-ray film; (2) data from this subject were free of obvious electronic artefacts, such as drop-out of one or more signals caused by problems with the recording apparatus; and (3) the data from this subject was the most temporally precise. With regard to point (3), it is important to note that the computer averaging of the utterances in Experiment #1 was aligned on the first vowel peak in the frame ([I] of "It's"). Öhman (1966) has pointed out that small variations in the length of the utterance will tend to influence the apparent amplitude of averaged emg signals. This is particularly true for peaks removed in time from the triggering point.

In examining the emg effects in the frame, we chose the 12 utterances containing the CVC syllables with initial and final /t/ and /k/ (with the vowels /i/ , /ɑ/ , /u/). Random fluctuation combined with certain well-known phonetic constraints (such as the fact that low vowels tend to be longer than high vowels, *ceteris paribus*) indicates that the length of

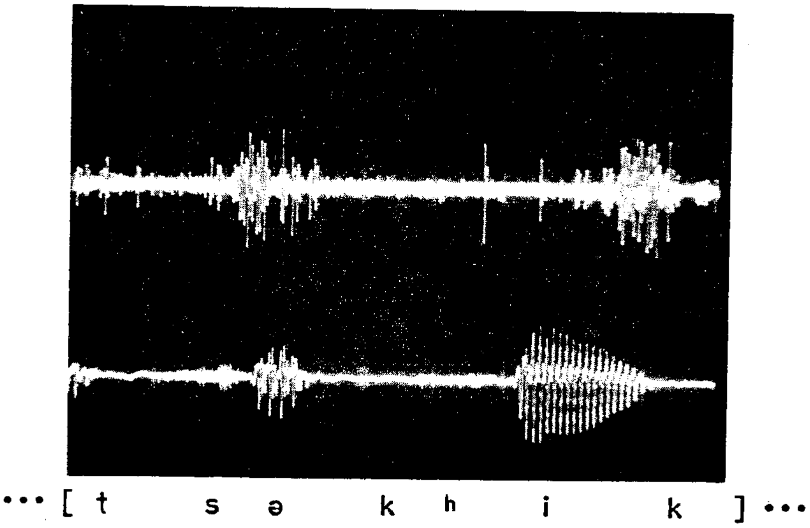


Figure 3.1

Single token of emg from the mylohyoid muscle during the utterance "It's a /kik/..." The time scale is 9mm = 50 msec.

these utterances will vary. For this subject, the variation was not very large. The time from the onset of the first vowel in the frame to the amplitude peak of the vowel [ə] in "again" was measured for each of the 12 averaged utterances. They ranged in length from 640 msec to 800 msec, mean of 740 msec, standard deviation of 44 msec. This is not an extreme range, and indicates that the amplitude of the later peaks in the utterance should not be too strongly influenced by the time distortion.

Table 3.1 presents the amplitude for each of the peaks of the mylohyoid muscle in the utterances containing the consonants /t/ and /k/ . The mylohyoid was active for the /ts/ of the first word (It's), the first consonant of the CVC syllable (C_1), the second consonant of the CVC syllable (C_2), and the /g/ in the last word (again). The amplitudes are given in arbitrary units on a linear scale (actually millimeters on the plotted output). One striking aspect of the data presented in this table is the variability of the activity for the first and last peaks, the /ts/ of "It's" and the /g/ of "again." Since these segments are each in similar phonetic contexts in the frame, the extent of the variation is rather remarkable. The mean amplitude for the /g/ peaks in column 5 of Table 3.1 is 37.2, standard deviation 8.4. The distribution of the amplitudes for the /ts/ peak in column 1 was even more widely scattered: mean = 19.5, standard deviation = 7.6.

Similar data for the posterior genioglossus muscle in the same subject is shown in Table 3.2. In this table, the figures in parentheses after the vowel peaks represent the averaged amount of activity of these vowels in the labial environment /p___p/ . Table 3.2 shows that the variability of the peak amplitudes for /ts/ and for /g/ was considerably less in the posterior genioglossus than in the mylohyoid. The mean amplitude for the g-peak is 21, standard deviation 3.2, and for the ts-peak the mean is 18.2, standard deviation 2.7.

The overall impression given by these data is that there is a fairly large amount of variability in the averaged emg data between utterances. In particular, the degree of variability found in the activity of the mylohyoid for stops in similar phonetic environments is quite surprising. It is conceivable that these gross variations in activity could be a function of an experimental artefact involving a long-term change in overall activity level, as might be caused by an effective change in the source impedance (the impedance of the muscle) resulting from a chemical interaction between the electrode tips and the muscle tissue, or by a slow drift in the gain of the amplifiers. This is unlikely, however, as the amplitude of the emg peaks in the phonetically similar contexts does not correlate with the order in which the items were recorded. Furthermore, there is no correlation between the amplitude of the audio signal for the item and the amplitude of the emg signal.

In order to determine the extent of the interaction among the emg peaks themselves a series of correlations (Pearson product moment) was

	1	2	3	4	5
Item	/ts/	C_1	V	C_2	/g/
tit	24	12	0	9	52
tɑt	18	14	0	16	28
tut	17	11	0	12	40
kik	17	36	9	38	37
kɑk	28	68	0	56	42
kuk	38	64	0	59	47
tik	13	8	4	53	38
tɑk	14	11	0	22	23
tuk	7	6	0	29	25
kit	16	51	9	6	42
kɑt	19	57	0	27	32
kut	23	51	0	22	40
Mean:	19.5				37.2
Standard deviation:	7.6				8.4

Table 3.1

Amplitudes (in arbitrary units) of the
averaged peaks in the mylohyoid muscle for subject TS

	1	2	3	4	5
Item	/ts/	C_1	V	C_2	/g/
tit	22	26	47 (55)	43	25
tɑt	15	18	0	27	17
tut	13	18	46 (53)	45	16
kik	17	35	57 (55)	46	18
kɑk	18	23	0	23	24
kuk	19	31	46 (53)	54	21
tik	21	27	56 (55)	51	22
tɑk	22	18	0	24	27
tuk	18	22	43 (53)	45	20
kit	17	28	52 (55)	19	18
kɑt	16	22	0	17	22
kut	21	26	53 (53)	23	23
Mean:	18.2				21
Standard deviation:	2.7				3.2

Table 3.2

Amplitudes (in arbitrary units) of the averaged peaks in the posterior genioglossus muscle for subject TS

	C_1	C_2	/g/
/ts/	.68	.42	.67
C_1		.99*	.39
C_2			.11 (.95**)

Table 3.3

Correlation data between peaks of mylohyoid activity (as represented in Table 3.1)

*indicates correlations where $C_1 = C_2$, and
**indicates correlation data for C_2 = /k/ (N=6)

	C_1	V	C_2	/g/
/ts/	.24	.21	.08	.84
C_1		.60	.57*	-.08
V			-.23	.00
C_2				-.21 (-.76**)

Table 3.4

Correlation data between peaks of posterior genioglossus activity (as represented in Table 3.2)
The vowel peaks are for /i/ and /u/ only.

*indicates correlation for $C_1 = C_2$ (N=6)
**indicates C_2 = /k/ (N=6)

performed on the data in Tables 3.1 and 3.2. These correlation scores are shown in Tables 3.3 and 3.4. The most noteworthy correlation scores in these tables deal with the effects between the first and last peaks of the utterances, and between the first and last peaks and the C_1 and C_2 peaks respectively. For the mylohyoid muscle (Table 3.3), there is a positive correlation between the amplitude of the peak for the /ts/ and the amplitude of the C_1 peaks, but also a high positive correlation between the /ts/ peak and the /g/ peak. The correlation between the C_2 peaks and the /g/ peaks is low (.11) if all C_2 are included (C_2 = /t/ or /k/), but nearly perfect (.95) where C_2 = /k/ is correlated with /g/ . The exact correlation between the C_1 and C_2 peaks where $C_1 = C_2$ is noteworthy. For the posterior genioglossus muscle, the first and last peaks showed an even higher correlation than in the mylohyoid, although the /ts/ peaks were not at all highly correlated with the C_1 peaks. In the posterior genioglossus, which is active in the production of high vowels and glides, the activity for high vowels seems to have an effect on the C peak, as is shown by the positive correlation (.60) between the vowels and the C_1 peak. The strangest result in the posterior genioglossus is the negative correlation between the C_2 peaks and the /g/ peaks. If all C_2 (/t/ and /k/) are compared to the /g/ , the correlation is negative, but low. If only the velars in the C_2 position are considered, a high negative correlation is found (-.76). These correlation data will be discussed in greater detail later on in this chapter.

The variability of the /ts/ and /g/ peaks, and the fact that these distributions are positively correlated, suggests that there may be an overall muscle activity level change from utterance to utterance. If this is the case, then it would be possible to correct for this factor in evaluating the differences in the C_1 and C_2 peaks. For example, each utterance can be considered with respect to the mean amplitude for the /ts/ or /g/ peaks. The difference between the mean and the observed amplitude can then be added to or subtracted from each peak for that muscle in that utterance. If this is done, however, the differences in C_1 and C_2 activity for different instances of the same phoneme are, on the whole, amplified rather than diminished.

The foregoing discussion clearly shows that the commands activating the mylohyoid and posterior genioglossus muscles in the production of the stop consonants /t/ and /k/ are not invariant for different intrinsic allophones of these phonemes. In fact, when this study was undertaken we did not expect to find such invariance. The process of speech production at the neuro-muscular level involves a fine and quite complex gradation of muscular activity. The level of activity present during the contraction of a muscle for a particular gesture in speech is a function not only of the immediate phonetic context but of a number of other factors. As mentioned above, factors such as degree of stress (Harris et al. 1969) are able to influence the motor patterns. Many other as yet undiscovered variables may play a role in the motor control of speech.

MacNeilage (1970), in a perceptive review article concerning the motor control of speech behavior, notes that there is every reason to expect that the motor behavior of the articulators in the vocal tract should vary considerably, depending upon, among other things, the previous state of the articulators. An ideal example of this was noted in Chapter 2 above, with regard to the behavior of the posterior genioglossus muscle in the production of the centralized vowel [ə]. If the centralized vowel follows a high or mid vowel, the posterior genioglossus is inhibited, and the centralized [ə] vowel is achieved by means of contraction of the extrinsic muscles which can lower the tongue body (the glossopharyngeus and, presumably, the hyoglossus). If, on the other hand, the [ə] follows a low vowel, some activity of the posterior genioglossus muscle is required to achieve the [ə] position. Furthermore, the [ə] segment is in no sense "phonemic" in English. It is a reduced vowel which can be the phonetic realization of a vowel which has not received sufficient stress in the phonological derivation of the utterance (Chomsky and Halle 1968). These arguments lead inevitably to the rejection of the assumption that the articulatory gestures in speech represent a variable response on the part of the articulators to invariant phoneme commands. MacNeilage (1970, p. 184) writes:

> ... it becomes clear that the more basic problem in speech production theory is not the one considered central to most theorists; namely, why articulators do not always reach the same position for a given phoneme. It is, How do articulators always come as close to reaching the same position as they do? One of the main conclusions of this paper is that the essence of the speech production process is not an inefficient response to invariant central signals, but an elegantly controlled variability of response to the demand for a relatively constant end.

As MacNeilage notes later on in his article, there are certain mechanisms involved in the control of skeletal muscle which could serve to modify central invariant commands so as to produce at least some of the variability found in muscle activity. In the following section we will examine critically the possible role of such mechanisms in the speech production process.

PERIPHERAL NEUROPHYSIOLOGICAL MECHANISMS AFFECTING THE CONTROL OF MUSCLES

By "peripheral mechanism" we are referring to any well-defined neurophysiological mechanism which does not involve direct cortical control, including the motor cortex and "psychomotor" cortical areas. There are, basically, only two such mechanisms: the so-called "gamma-efferent" or "gamma-loop" mechanism; and the mechanism that we shall call, following Eccles (1969), the "cerebellar loop." The gamma-efferent system will be discussed at length, since at first sight it appears to provide a viable system for modifying the muscular gestures of speech to make them compatible with previous states of the vocal tract. Cooper (1966), Ladefoged and Fromkin (1968), Campbell (1968), Öhman (1967b), MacNeilage and DeClerk (1969), Tatham (1969a), and MacNeilage (1970) have, among others, suggested that this mechanism might be involved in the control of the muscles of speech.

The gamma system*

Skeletal muscle in mammals contain small proprioceptors known as "muscle spindles." These are found in great profusion in spinal muscle, and are also to be found, in lesser numbers, in most but not all cranial muscles. The spindles serve several purposes. Their main purpose is to relay information concerning the length of the muscle and the rate at which the length is changing back to the spinal cord or the brain stem. Some of this information is projected to the cerebellum (Eccles 1966). This information serves to modify the activity of muscles in various ways. The spindles are small fusiform bodies located in parallel with the extrafusal muscle fibers. Within a spindle are found two types of intrafusal fibers, nuclear bag fibers and nuclear chain fibers. The distinction between these fiber types is not important for purposes of this discussion, but it must be noted that both types of fibers are, at least in part, muscular, and receive motor innervation. Figure 3.2 is a detailed drawing of the internal part of a typical mammalian muscle spindle. Note that the intrafusal fibers receive motor innervation from the thin γ-efferent fibers, and afferent (sensory) innervation from the thicker Type Ia and Type II fibers. The Ia and the II fibers have different terminations on the spindle fibers. The unmyelinated end of the Ia fiber loops around the central portion of the spindle fibers (annulospiral ending), and the Type II fiber ends in a so-called "flower-spray" configuration. These endings are now usually called primary (Ia) and secondary (II) endings. Both the primary and secondary endings are able to sense the instantaneous length of the spindle, and relay this information back to the central nervous system. In addition, the primary endings can sense the velocity at which the length of the spindle is being changed (Matthews 1964). Sensory information from the muscle spindles does not project to the cortical areas (Cooper 1960). This is the basis for the claim that the muscle spindles are not involved in conscious sensation of the position of various parts of the body (Matthews 1964).

---

* In the brief discussion below, technical details and complications are kept to a minimum. The best general introduction to muscle receptor anatomy and function is in Eyzaguirre (1969, Chapter 7). For more specific information, consult Cooper (1960), Matthews (1964), and Granit (1955, 1966 and 1970). The integration of the gamma system within the extra-pyramidal motor system is discussed, somewhat briefly, by Jung and Hassler (1960).

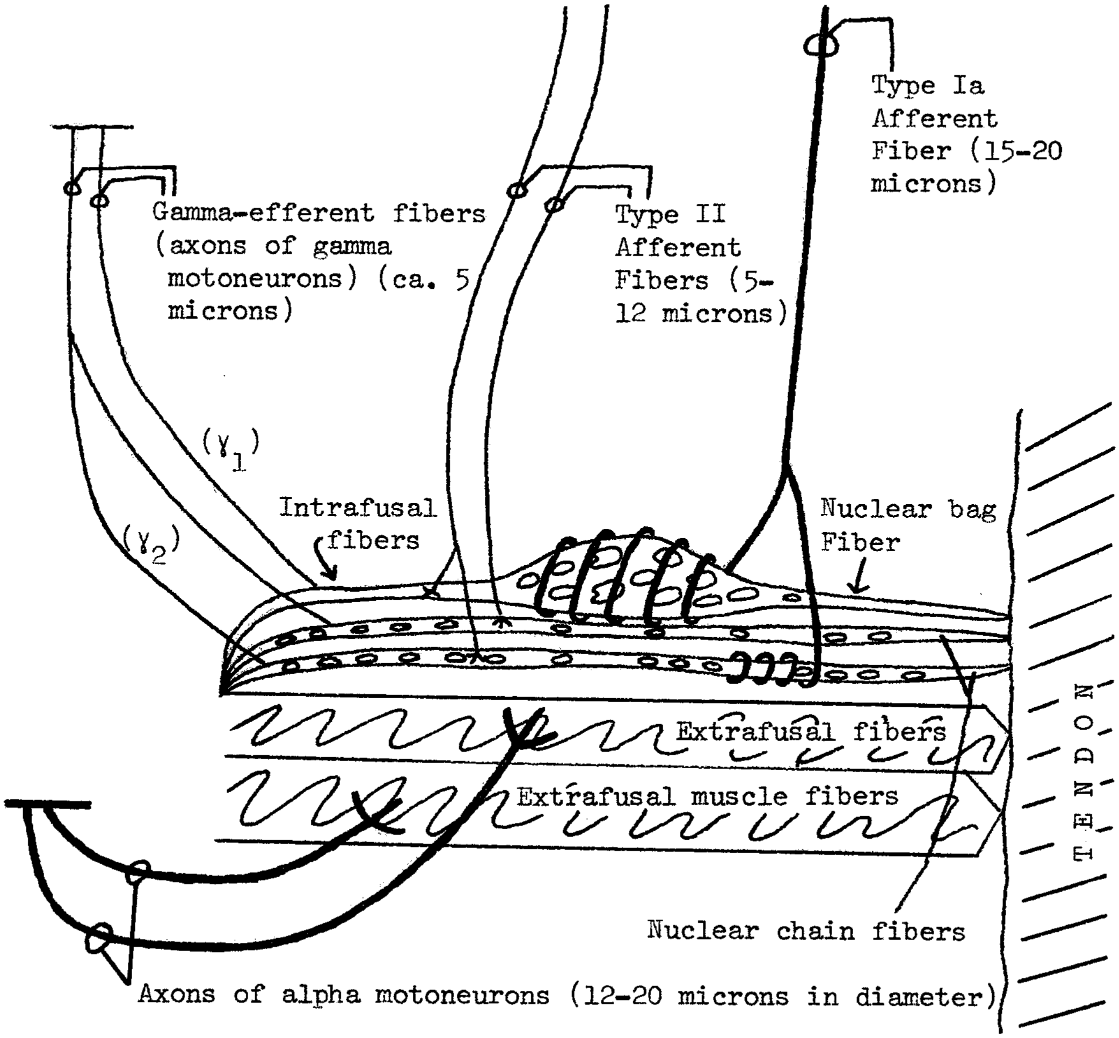


Figure 3.2

The motor and sensory innervation of a typical mammalian muscle spindle. The extrafusal fibers are part of the muscle itself. The intrafusal fibers are the contractile part of the spindle, and are innervated (motor) by the $gamma_1$ and $gamma_2$ fibers. Drawing not to scale.

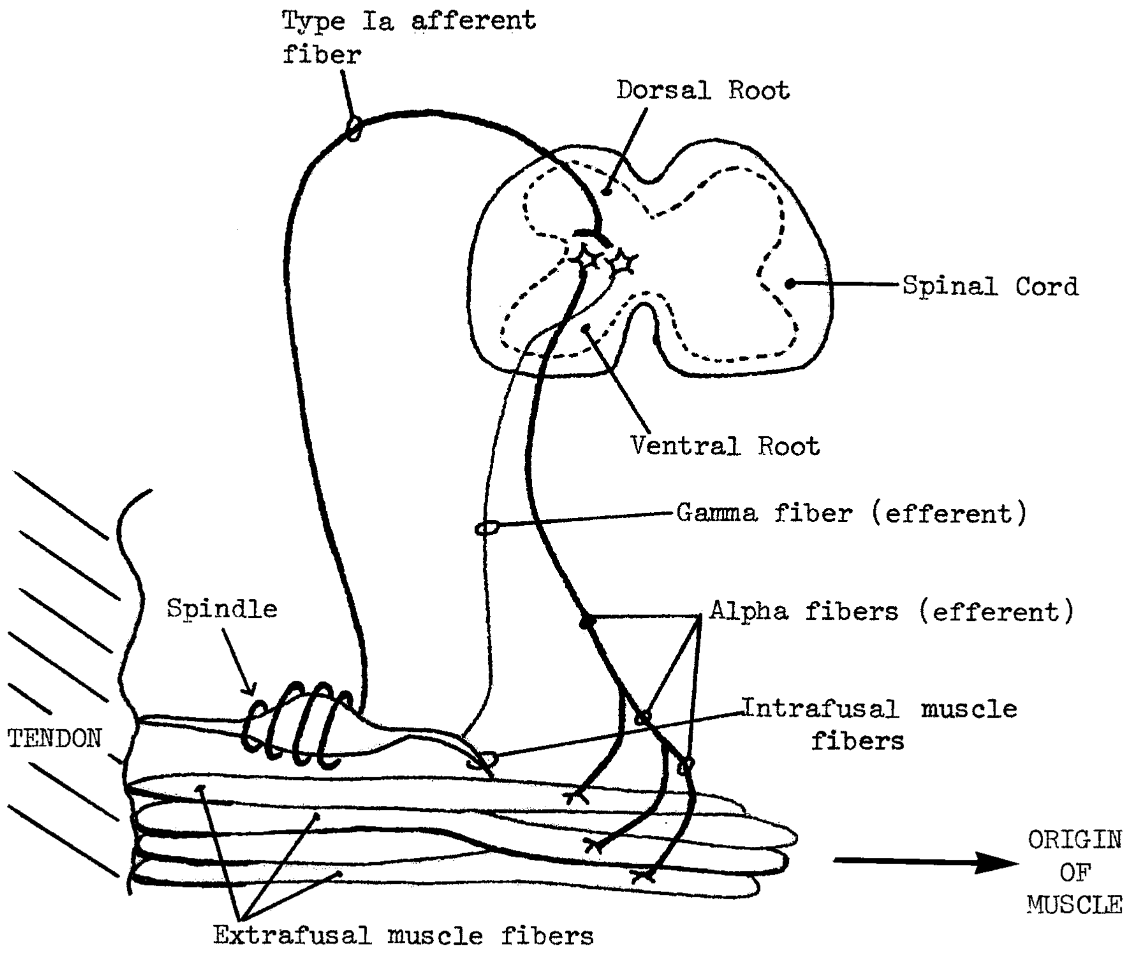


Figure 3.3

Simplified view of a muscle spindle, showing its connections to extrafusal muscle fibers and to the spinal cord. The afferent fiber from the spindle forms a synapse in the spinal cord on the alpha efferent fiber which innervates the extrafusal muscle fibers. Drawing not to scale.

Extrafusal muscle fibers receive motor innervation from α motoneurons, which have large diameter fibers (12-20 microns, typically). These are also shown in Figure 3.2. The mechanical and neuronal connections between a spindle and both the extrafusal fibers and the spinal cord are shown in Figure 3.3. In this illustration only a Type Ia sensory fiber is shown forming a monosynaptic arc with the α motoneuron. It is important to realize that this illustration represents a considerably simplified view of the muscle receptor system. Both Type Ia and Type II fibers can affect the firing of α motoneurons, in fairly complex ways.

The muscle spindle afferent neurons synapse on α motoneurons in the spinal cord. By this means, when the spindles in a muscle are stretched, they can reflexly control the firing of the α motoneurons. This control can take the form of inhibition of α motoneuron pools of antagonistic muscles through inhibitory interneurons as well as activation of the α motoneuron pools of synergistic muscles. This is the lowest-level basis of the mechanism of reciprocal innervation of agonist and antagonist muscles.

In order to evaluate the possible role of the gamma system in speech, it is necessary to discuss further the ways in which the gamma system can act in the control of skeletal muscle. Purely for the sake of discussion, the gamma system can be thought of as having three possible modes of activity:

A.) In the "stretch reflex" mode, the muscle spindles would respond quickly to any externally-caused change in muscle length, and, by firing the α motoneurons, would cause the muscle length to be quickly restored. In this mode, the gamma motor fibers apparently play no role. This is the basic mode of action of the myotatic reflex.

B.) In the "muscle biasing" mode, the γ-*efferent* system fires, causing the intrafusal muscle fibers to contract. This action stretches the spindle, firing the Ia and II afferent fibers, thus reflexly firing the α motoneurons. The γ-fiber firing is apparently originated in the brain stem, and can be thought of as setting a base level of contraction for the muscle. This "bias" level is maintained by α motoneurons fired monosynaptically by the spindle afferent fibers. This mode of gamma activity, upon which ordinary α-efferent activity under higher-level control is superposed, can be thought of as analogous to a DC balance control in an amplifier. It sets up a basic tone or "posture" for a muscle, independent of the corticospinal (pyramidal) motor system.

C.) The third possible mode of activity of the gamma system could be termed the "gamma-drive" mode. This mode is similar to B in that the gamma efferent fibers contract the spindle fibers, but in

> this case they initiate the contraction of the muscle by setting the spindle to a certain length. The α motoneurons then fire monosynaptically, contracting the muscle to a fixed length which is proportional to the length of the spindle fibers. This mode of operation is of considerable interest for models of speech production, since it is conceivable that it could be a mechanism by which a muscle can be set to a desired "target" length by one invariant command. The actual amount of α motoneuron activity would depend upon the previous state of the muscle.

Modes B and C are of most interest to models of speech production. The basic difference between them is that in mode B both the γ and α motor systems are firing, and the γ-system is serving to maintain a basic muscle posture, while in mode C the γ-system becomes primary, in a sense, being the system through which the muscle contraction is initiated. It has been known for a long time that both γ- and α-efferent activity contributes to the movement of skeletal muscle (Granit, Holmgren, and Merton 1955). The role of the γ system in reflex activity, and its role in the initiation and control of tonic muscle contraction is well established. The role of the γ system in phasic muscle activity of the sort employed in speech is less clearly established. The question is complicated by the fact that most of the detailed studies of the gamma system have been performed on the spinal muscles of non-primates. In dealing with the activity of cranial muscles in man we are in a veritable physiological "no-man's land." A careful search of the literature reveals that there have been no neurophysiological studies of possible gamma system activity in human cranial muscles. This is not surprising, since such experiments often involve manipulations which are ultimately fatal.

Some studies have been done on the latency of muscle control in the spinal muscles of man. Campbell (1968) speculated that muscle spindles were involved in the control of the respiratory muscles. Sears and Newsom Davis (1968) recorded electromyographic signals from the intercostal muscles during respiration, as well as measures of respiratory activity such as airflow and lung volume. They demonstrated that muscle responses to loading and unloading of the respiratory system were of relatively short latency (33-80 msec), indicating that these responses are not voluntary. The latencies observed were claimed to be consistent with the transmission time around a gamma loop. Nevertheless, as MacNeilage (1970) has pointed out, these results indicate only that the possibility for such a control system exists, and that it can be brought into play in abnormal circumstances. They do not conclusively demonstrate that the on-going control of speech is handled by such mechanisms. Furthermore, the action of spinal muscles such as the intercostals is relatively slow compared with the fast actions of the larynx, tongue, and lips (Draper, Ladefoged, and Whitteridge 1960). Much less temporal precision is required in the control of the respiratory muscles in speech than is required for control of tongue position.

In discussing models of motor function incorporating the gamma system, it will be necessary to adduce some evidence based upon histological and physiological studies of several non-human species. As we have mentioned before, direct evidence from human experimentation is scanty. It is necessary to keep in mind that what applies to the cat or the monkey may not hold true in man, and that some of the arguments must be considered to be tentative, at least to the extent that they are based on such evidence. Nevertheless, there are sufficient basic similarities in the motor systems of most mammals to make the arguments at least indicative.

The muscles employed in speech are innervated primarily by the cranial nerves. The tongue muscles with which we are concerned are all innervated by the XIIth cranial nerve (the hypoglossal), with the exception of the mylohyoid. This muscle receives motor innervation from the mandibular branch of the Vth cranial nerve (the trigeminal). As was mentioned earlier, the cranial muscles are not so liberally supplied with spindles as are the spinal muscles. However, muscle spindles have been found in the intrinsic muscles of the larynx (Lucas Keene 1961), and in the intrinsic tongue muscles of man and cats (Cooper 1953). Cooper also reported seeing spindles in the genioglossus muscle. Voss (1956) reported spindles in the human geniohyoid, a muscle innervated by the hypoglossal nerve, however he found no spindles in the mylohyoid. In cranial muscle, the number of spindles per gram of muscle weight is usually around 5, while a figure of 20 is more common for the spinal muscles (Cooper 1960, quoting from Voss 1956).

One major problem in evaluating the role that spindles could play in the control of the tongue muscles is that the pathway of the spindle afferents is not definitely known. There has been a considerable degree of controversy concerning the status of the hypoglossal nerve. Anatomy texts normally consider the nerve to be purely motor, containing no sensory fibers. Cooper (1953) reported that she was able to record afferent signals produced by tongue stretch from the hypoglossal of the cat. These findings have not, in general, been confirmed (v. Porter 1965). Stimulation of the hypoglossal nerve trunk produces only antidromic spikes, and not the afferent signals typically associated with Type Ia and Type II sensory spindle fibers. Furthermore, caliber spectra of this nerve reveal that the distribution of fiber diameters is unimodal, which is usually not the case in a combined motor-sensory nerve (Blom 1960). Blom concluded that the hypoglossal nerve in both man and cats is motor only. Whatever the eventual resolution of this controversy may be, it is clear that the hypoglossal does not contain large Ia or II afferent fibers. Smaller diameter afferent fibers could exist, possibly fibers innervating other non-spindle proprioceptors (Law 1954).

Stimulation of the hypoglossal nerve does produce afferent signals in the lingual nerve (a branch of the Vth, or trigeminal nerve), however.

Szentágothai (1948) described collaterals of afferent fibers of cells in the mesencephalic trigeminal nucleus which projected to the hypoglossal nucleus, however these afferents were sensory to muscle spindles in the masseter and temporalis, muscles of mastication. Porter (1965) detected low-threshold afferents in the lingual nerve of cats which produced synaptic effects on hypoglossal α motoneurons, however he considered these to be impulses from exteroceptors (mechanical receptors) in the tongue, not muscle spindles. In general it seems unlikely that there are extensive monosynaptic connections between the sensory nerve fibers which project to the trigeminal nucleus and the hypoglossal motor neurons. A large number of such monosynaptic connections would have to be demonstrated in order to establish the possibility of follow-up servo-control of the tongue muscles using a monosynaptic loop.

It is noteworthy that certain activities, such as licking and swallowing, involving the tongue are maintained even in decerebrate cats (Porter 1966), and basic innate patterns such as yawning, sucking, and swallowing can be carried out by anencephalic children (Jung and Hassler 1960). Nevertheless, the fact that basic patterns such as swallowing and sucking can be totally under non-cortical control does not have any significant implications for models involved in the control of speech behavior.

In summary, the following points relevant to models of gamma system control of speech have been noted:

1.) The distribution of spindle-type proprioceptors in the cranial muscles is more limited than in the spinal muscles.

2.) Muscle spindles have, however, been found in the tongue muscles.

3.) The hypoglossal nerve does not contain afferent fibers of the type normally associated with muscle spindles.

4.) Afferent activity from the tongue, quite possibly including spindle afferents, is carried by the lingual nerve.

5.) There are connections between these nerve nuclei in the brain stem, however there is not likely to be a large field of monosynaptic projections.

The question of whether the gamma system is used in speech is much like the more basic question of whether feedback in general is used in speech. The question is not *whether* the gamma system is used, but *how* it is used. In some discussions of gamma system activity in speech, it is proposed that the system could act in such a way as to preserve the notion of segment invariance at a slightly more abstract level. This position was taken by Ladefoged and Fromkin (1968), and is specifically stated as an hypothesis by Öhman (1967b), Tatham (1969a), and MacNeilage and DeClerk (1969). MacNeilage and DeClerk propose that the gamma system would provide an explanation for many of the left-to-right context effects found in their study. They found that the electromyographic activity at many electrode locations on the tongue and facial muscles in their study (surface electrodes were used) differed for a particular phoneme depending upon the (phonetic) nature of the preceding phoneme. They studied CVC syllables uttered in isolation, and found that the first consonant influenced the emg activity for the vowel, and that the nature of the vowel had a marked influence on the activity for the following consonant. These effects are not very surprising, as they noted, since the amount of contraction required for a muscle to produce a segment (S_n) will depend upon the length of that muscle during the articulation of the preceding segment (S_{n-1}). If, however, the commands to the muscles are initiated by the gamma efferent fibers, an invariant gamma command for a segment (S_n), setting up an essentially invariant degree of stretch on the muscle spindles, would result in a variable degree of α motoneuron activity, the activity being directly proportional to the muscle length for segment (S_{n-1}).

There are several problems with this claim. First, MacNeilage and DeClerk themselves noted that there are data in their study which cannot be handled by such a mechanism. For example they found that the elevation of the hyoid bone is 3-3 1/2 mm. less for final /g/ following high vowels (/i/ , /u/) than following low vowels (/æ/ , /ɔ/). They inferred from this that "the length of muscles controlling the position of the hyoid is not constant in all cases of final /g/ , even though in utterance final position there are no further actions to interfere with the attainment of a constant hyoid position" (p. 1232). In this case, it is conceivable that the lower hyoid position could be a result of increased sternohyoid activity. The sternohyoid, in effect a depressor muscle of the larynx, is active both for voiced stops (where the larynx is lowered) and in final position in an utterance (Hirano et al. 1967). Since the hyoid is higher during the vowels /i/ and /u/ than during /æ/ and /ɔ/ , it is possible that a large amount of sternohyoid activity would be found in the utterance-final position in this case (MacNeilage and DeClerk did not use a frame for the CVC utterances). There is no direct evidence for this, however.

There is a similar but more conclusive piece of evidence counter to gamma control in the present study. We found that the mylohyoid activity is consistently greater for a /k/ preceding and following an /ɑ/ than for the /k/ in the environments _ i _ and _ u _ . Furthermore, we found that the hyoid is more elevated throughout the whole utterance containing the item /kɑk/ than the utterance containing /kik/ . This is shown in Figure 3.4. If we assume that the height of the hyoid bone is an approximate indication of the length of the mylohyoid muscle, then the mylohyoid muscle is from 2-3 mm. *shorter* during the vowel /ɑ/ in these syllables than during the /i/ . During the closure for the final

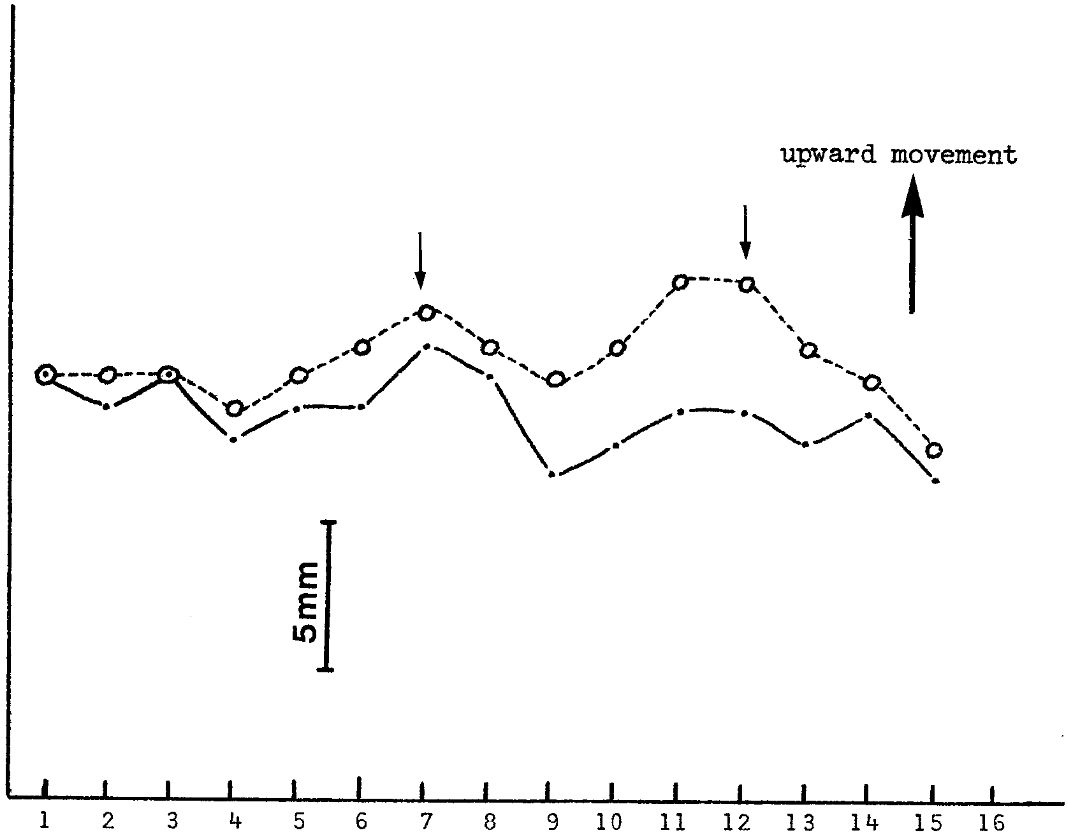


Figure 3.4

Movement of the anterior part of the hyoid bone in a vertical plane. The dotted line represents hyoid elevation for the utterance "It's a /kak/..." and the solid line elevation for the utterance "It's a /kik/..." The small arrows represent the approximate points of closure for the velar stops in the CVC words.

velar stops (indicated by the second arrow on Figure 3.4), the mylohyoid is about 4 mm longer during the velar stop following /i/ than during the /k/ following /ɑ/. The mylohyoid is almost as long during the final /k/ after an /i/ than during the vowel /ɑ/. This situation, which is represented schematically in Figure 3.5 above, could not be the result of an identical gamma-drive command to the mylohyoid for both intrinsic allophones of the final /k/. This observation, coupled with the fact that the mylohyoid is not found to contain muscle spindles, indicates that the systematic variations in the activity of the mylohyoid muscle for the intrinsic allophones of the final /k/ in the syllables /kik/ and /kak/ cannot be accounted for by a gamma-drive model of muscle control. As we will suggest in the next chapter, it is probable that the activity of the mylohyoid muscle is regulated by the degree of mandible opening. Syllables like /kak/ appear to have an overall greater degree of jaw opening than syllables like /kik/. If this is indeed the case, the activity of the mylohyoid muscle would be increased during the whole syllable in order to provide for dorsal stop closure, which would account for the increase in height of the hyoid bone during the whole syllable.

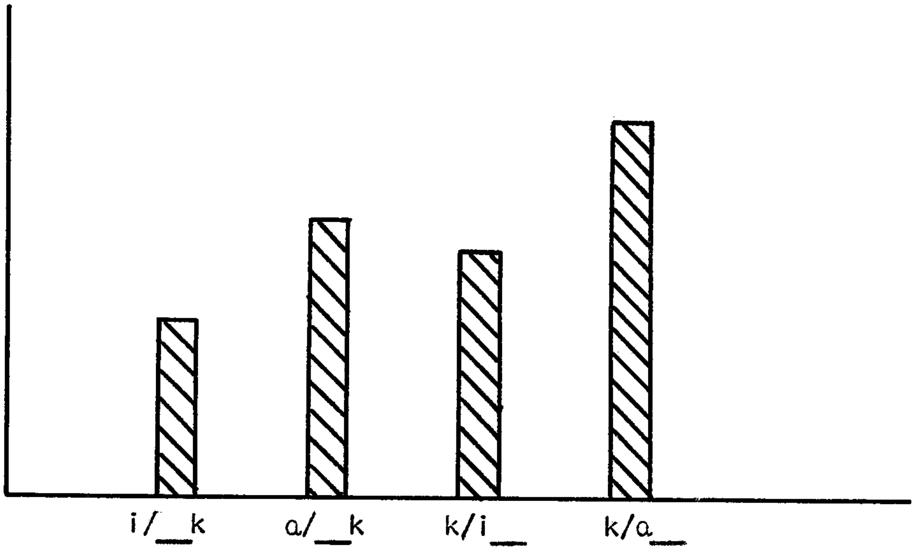


Figure 3.5

Schematic representation of the height of the hyoid bone during /i/ and /ɑ/ and during /k/ following /i/ and /ɑ/.

There are several general objections to the postulation of an invariant gamma-drive model producing the variable activity of the tongue muscles. The concept of a short latency monosynaptic reflex arc is not necessarily applicable to the tongue muscles innervated by the hypoglossal nerve, for the reasons noted above. It is quite possible that the gamma system may play a role in the control of the cranial muscles in speech, however it is not likely that it operates as a simple servo, setting up a (monosynaptic) length follow-up loop with the α motoneurons. Additional negative evidence is provided by consideration of the relative fiber diameters of the γ and α neurons. The γ-efferent fibers are rather small, having an average diameter of 5μ or less. This, in turn, means that their conduction times are slow, on the order of 30 meter/sec. or less (Eyzaguirre 1969). Also, some of the smaller γ_2 fibers may be unmyelinated (Barker 1966), which would further decrease their conduction speeds. Alpha fibers typically have conduction speeds of around 100 meters/sec. MacNeilage (1970) has observed that segment commands in running speech "would require different sets of commands being issued by the gamma motor system about every 70 milliseconds." He further notes that "Such rapidly varying phasic control has not in the past been associated with the gamma system" (p. 191). It seems doubtful that fibers having such slow conduction speeds would be centrally involved in the production of speech. It was noted above that the tongue muscles are among the fastest in the human body, exceeded only by the muscles of the eye. To suppose that their activity is under the principle control of the essentially slow gamma motor fibers would negate much of the advantage for speech provided by this speed and precision.

Other investigators have taken a slightly different view of the role of the gamma system in speech production. Öhman (1967b) has postulated that the gamma system serves to "accelerate the firing of motoneurons during the initial phase of a central excitation command" (p. 41). The argument is presented as follows:

> For many muscular functions a central excitatory command is usually distributed to most of the motoneurons that innervate the muscle, including the motoneurons of the muscle spindles. Since the spindle motoneurons are among the smallest fibers contained in the efferent nerve they will be the earliest in responding to the excitatory command. Consequently, the muscle spindles will start contracting earlier than the majority of the extrafusal fibers.
>
> Since the contraction of the extrafusal fibers initially lags behind that of the spindles, the receptor organs of the latter will be activated owing to stretch

> and a train of afferent action potentials will be transmitted to the motoneurons of the extrafusal fibers. In the somata and dendrites of the latter an excitatory synaptic current will thus be added to that produced by the central direct command. As a result of these events the motoneuron firing frequency becomes higher and the proportion of activated fast fibers greater than it would be without the gamma loop.
>
> After a while, however, when the contraction of the extrafusal fibers has caught up with that of the spindles, the afferent discharge of the latter decreases. In this phase of the contraction, the motoneuron and muscle fiber activity therefore settles down to a lower steady value and becomes more directly related to the intensity of the central command. (p. 41-42)

Öhman concludes that this mode of operation could account for the peripheral variability in emg signals of the facial muscles, in that the α motoneuron discharge could be *increased* (or decreased, presumably) by spindle discharge. The spindle discharge in his model would vary with the previous state of the muscle.

> A central command of constant intensity delivered to the orbicularis muscles as for the lip rounding gesture of an [y], for example, should produce an onset EMG burst of quite variable amplitude depending on how spread or rounded the lips are when the gesture starts. The more spread the lips are initially, the greater will the stretching of the orbicular muscle spindles be, and the greater the command amplifying the activity of the afferent [sic] neurons. (p. 42)

It is true that, for a given group of neuron somata, the smaller cells will respond earlier, since they will fire at a lower threshold (Henneman, Somjen, and Carpenter 1965). Nevertheless, these differences in response time are quite small (several milliseconds or less), and this effect would certainly be overcome by the fact that the impulses will travel considerably more slowly over the axons of the smaller motoneurons. This is because, as mentioned above, the speed of propagation of a nerve action potential depends, among other things, on the diameter of the axon. A propagation speed difference of 70 meters/sec. can lead to a 3-4 msec difference in transmission time over the axon lengths innervating the speech muscles.

We conclude that it seems unlikely that a pure "gamma-drive" model can be postulated to account, at a peripheral level, for the variability of muscle motor unit activity. That is to say, the gamma system *by itself* is inadequate to account for the variability of muscle activity caused by the immediately preceding state of the muscles of the articulators. The gamma system is poorly adapted as a primary drive mechanism for the fast muscles involved in the precise and rapidly-adapting activity characteristic of speech. This conclusion is totally consistent with the vast body of currently available physiological evidence. Matthews (1964) writes:

> In discussing the γ route of excitation it is necessary to keep somewhat apart the evidence which indicates that the length of a muscle is maintained at a certain value by means of a servomechanism, and the evidence which suggests additionally that the null or balance point of the servo is normally changed by fusimotor [gamma efferent] activity acting alone. The evidence for the former idea is considerably stronger than that for the latter. Indeed, on present evidence it could be suggested that the main function of the fusimotor fibers is to control various parameters of the servo (stiffness, damping) rather than to alter its balance point by biasing the spindle. (p. 277)

Furthermore, Matthews notes the possibility that "the distribution of Ia fibers from the primary endings of some muscles is so diffuse, to so many kinds of α motoneuron, that the γ loop could not be used to control many fine movements" (p. 276). The gamma system does appear to be able to act, in some cases, to correct for over- or under-excitation by the α motoneurons. Again Matthews:

> It has been suggested as a feature of the "γ loop" that this route alone permits the valuable servo properties of the stretch reflex to be maintained during shortening. But this is not necessarily so, as in many servomechanisms activity can be injected into several different parts of the loop with rather similar effects. In the present context this means that activation of muscle via the α route can maintain the advantages of a servo, provided that the shortening does not occur so rapidly that the spindle endings cease discharging altogether. Of course, the reduction of spindle discharge caused by the shortening will reflexly oppose the direct excitation of the motoneurons by higher centers, so that any direct excitation by the "α route" must be made appropriately powerful. The α route would perhaps be most efficiently employed in conjunction with sufficient fusimotor activity to prevent any decrease in spindle discharge occurring during the contraction; this would be achieved if the relative amounts of α and of γ activity were adjusted to be appropriate for the velocity of shortening "expected" under any particular set of conditions. Then

> if shortening proceeded faster than "intended" by the higher centers it would be slowed by servo action, and if shortening were hindered by some unexpected load it would be speeded up by servo action. Such a mode of action would not suffer from the slowness inherent in the excitation of muscle by the γ route alone, and would agree with the experimental finding that fusimotor neurons and ordinary motor neurons are often activated together, even when the dorsal roots are cut. (p. 277)

It is very tempting to misinterpret the last part of this quotation. The type of modifications that the γ system can provide, as described in the quote above, are probably used in such activities as maintaining an upright stance in a rocking boat, or walking over uneven ground. Drastic overshoot and "unexpected loads" are not, however, typical in normal speech. When they are introduced, as in attempting to speak with a clenched jaw, or speaking with obstructions in the mouth, it is possible that the gamma system may play a role in readjusting the action of the articulators. The role of the gamma system in normal speech is still in doubt. One of the most widely noted characteristics of normal speech is its temporal precision. Although the gamma system is the most rapidly acting mechanism in the nervous system, it can still take from 30-50 msec for the gamma system modifications to have an effect. (The actual neuron and synapse transmission times are quite fast, less than 10 msec. However the response times of the spindle and muscle fibers must also be included.) Even this rapidly acting loop could be too slow for gestures which can, under fast rates of speech, be made every 70 msec or even faster.

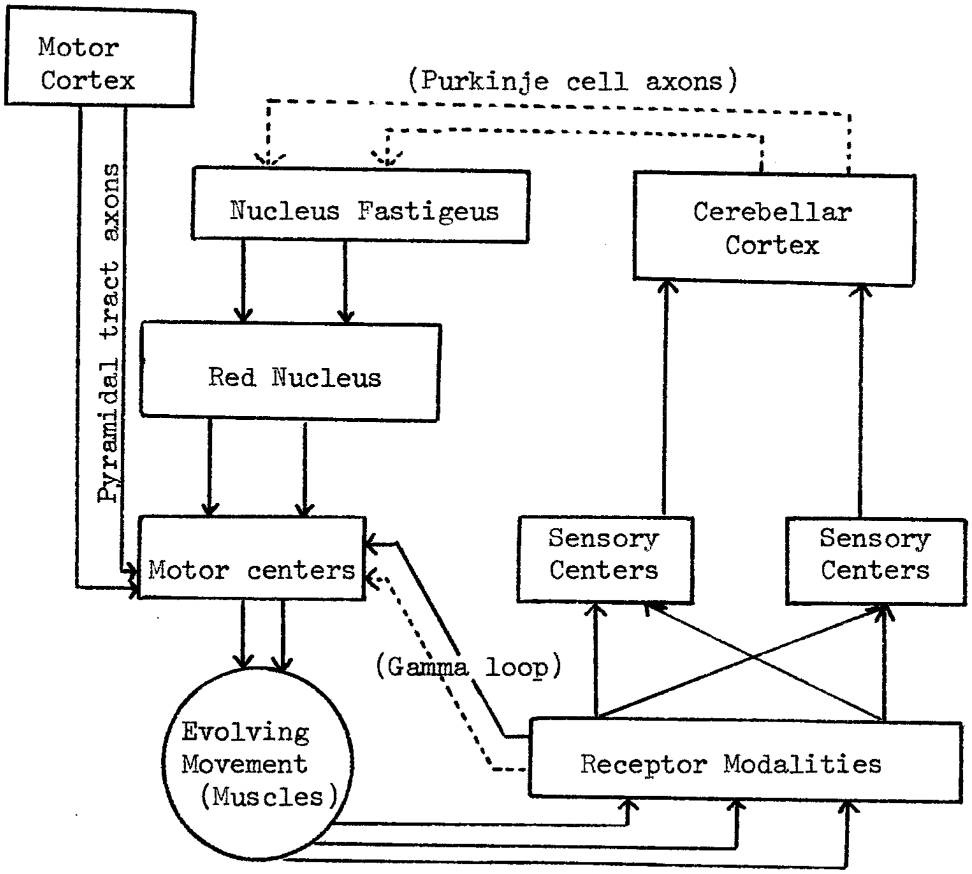
In summary, it can be stated that the gamma efferent system most probably does not act to *initiate* articulatory gestures. It may play a role in *correcting* ongoing gestures, as suggested by Öhman (1967b), although there is no solid evidence that it does.

Cerebello-muscular mechanisms

The cerebellum plays an important role in the regulation and coordination of motor activity (Eccles 1966, 1967; Eccles, Ito, and Szentágothai 1967). Cerebellar disorders in man and other mammals tends to produce lack of coordination of movement, and induces errors in force and rate of movement. Decerebellate animals normally evidence *ataxia*, a disorder of normal gait in which correct adjustment of limb movements cannot be made. It is clear that the cerebellum plays an important function in spinal motor control. Its integrative function in the specific motor activity of speech is not so well documented. Lesions of the cerebellar cortex do produce disturbances in speech, particularly disarthric disturbances which typically show a lack of temporal coordination between the various muscles used in speech production.

Recent developments in neurophysiology have led to the postulation of models of cerebellar function which demonstrate how the cerebellum can serve to control and modify muscle activity. The cerebellum receives sensory information from the muscles (spindle afferents) and from other sensory receptors in the body (joint receptors, tendon organs, tactile receptors of various sorts, etc.). The cerebellum can influence the control of muscle through interneurons in the reticular nucleus and Deiter's nucleus. Figure 3.6 demonstrates the so-called "cerebellar loop," which may be used to modify commands from the motor cortex. In this model, the motor commands from the cortex are transmitted to the muscles via the large and small pyramidal tract neurons. Various parameters of the resulting muscle action (force of contraction, muscle length, and rate of change of muscle length) are sensed by the receptor modalities (including, but not limited to, the muscle spindles), and this information reaches the inhibitory neurons of the cerebellar cortex (the Purkinje cells). The Purkinje cell axons synapse on interneurons in the reticular nucleus (and in Deiter's nucleus), the output of which can inhibit or excite the α motoneurons in the brain stem (or in the spinal cord). This control loop is conceptually similar to the gamma loop. It is however, not monosynaptic, and is clearly much more sophisticated in function. It can sense and integrate a much wider range of information than the lower-level gamma-loop, and can influence the control of muscles which are widely separated in terms of both location and function, rather than simply agonist-antagonist groups as is the case with the gamma loop. It is quite possible that basic innate patterns like swallowing, vomiting, yawning, etc., which involve precise temporal integration of a large number of muscles, are initiated and carried out totally within a loop such as this. Notice that the gamma loop could form a sub-loop within this mechanism, as shown in Figure 3.6.

What role could the cerebellar loop play in the control of the gestures of speech? At first sight, it seems possible that the commands to the muscles provided by the pyramidal tract axons could be modified by the cerebellar loop, thus providing a means by which invariant commands to the muscles could result in an extremely variable degree of muscle activity. This simple hypothesis becomes much less attractive when the cerebellar processing times are considered, however. Eccles (1969) has estimated that the transmission time around the cerebellar loop in the control of the human arm muscles would be on the order of 100-110 msec. For the muscles of speech, the transmission time would be reduced somewhat, due to the shorter lengths of the motor and afferent nerve axons involved, as well as the faster twitch times of the speech muscles. However the transmission time around this loop is not likely to be less than 80-90 msec. Given these time constants, an invariant command sent *via* the pyramidal tract could not be made compatible with the previous state of the vocal tract at all fast enough to correct the articulatory gestures. In this study, bursts of emg activity frequently lasted no more



Excitatory path ———

Inhibitory path --------

Figure 3.6

Schematic diagram showing the function of the cerebellar loop in the control of evolving movement. The gamma loop has also been included. Adapted and simplified from Eccles (1969).

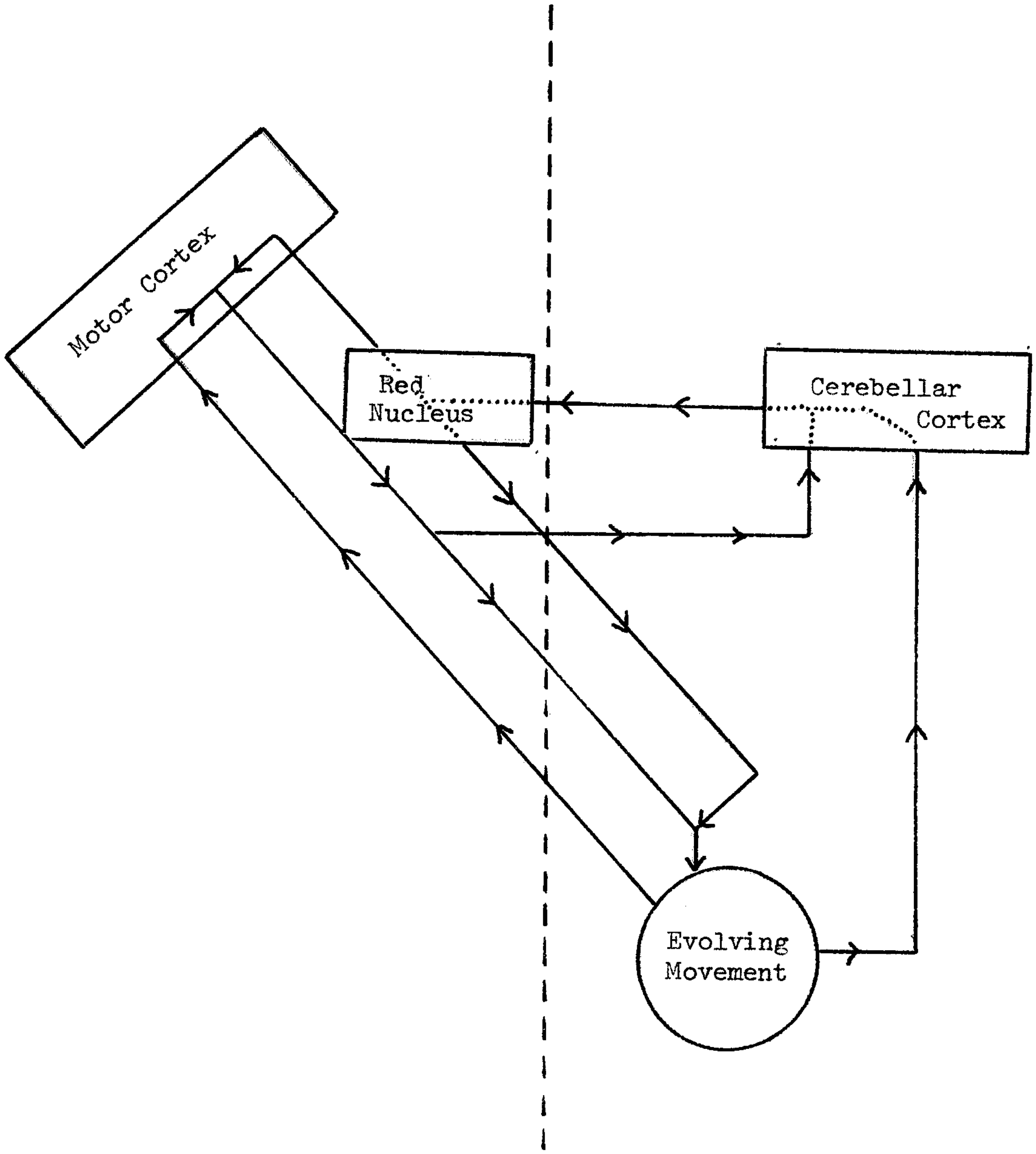


Figure 3.7

Schematic diagram showing the interconnections between the cerebellum, reticular nucleus, and the motor cortex. Broken line represents the midline, showing the decussation of the fibers. Adapted from Eccles (1969).

than 80-100 msec. This is particularly noticeable in the activity of the posterior genioglossus and mylohyoid muscles for the articulation of the velar stops. Given these facts, it is highly unlikely that the cerebellar loop by itself would play a major role in reorganizing the motor commands of speech. This argument applies only to the possibility of sub-cortical reorganization of the motor commands by the cerebellar loop. We shall see later that the sensory information on the state of the motor system provided by the cerebellum could possibly play a role in the reorganization of commands to the muscles prior to their leaving the cortical areas.

In summary, it is clear that the demands made upon the neuro-muscular system by speech activity are particularly severe. Commands must be delivered to more than 20 muscles at extremely fast rates, and these commands must be temporally and spatially integrated in highly complex ways. The exact methods by which either of these types of integration is achieved are not known, even in the most generalized way. It is likely that the peripheral neuro-muscular mechanisms play a substantial role in both the temporal and spatial reorganization of the muscle activity. Nevertheless, it is important to realize that these mechanisms alone apparently cannot provide all of the reorganization needed in these complex processes. In particular, it is felt that the role that the gamma-efferent system could play in the reorganization of speech gestures has been vastly overemphasized in the speech production models that have been proposed so far.

CORTICO-CEREBELLAR CONNECTIONS

In addition to the connections with the muscle receptors and alpha motoneuron pools, the cerebellum also has rather extensive connections with the motor cortex and the association or "psycho-motor" areas of the cortex (Jung and Hassler 1960; Eccles 1969). The cerebellar hemispheres and intermediate lobes have outputs which pass through the thalamus and various brain stem nuclei such as the reticular nucleus, and can affect the output of the pyramidal tract neurons. This situation is illustrated schematically in Figure 3.7. These connections serve to keep the cerebrum informed of the nature of the evolving movement, and allow certain modifications to be made to the muscle commands before these commands are issued from the motor cortex.

The role which such connections could play in reorganizing the motor commands to the muscles has been discussed by MacNeilage and DeClerk (1969). In particular, they propose two mechanisms which might, using these cortico-cerebellar connections, explain part of the context effects so often noted in emg data: an "anticipatory" mechanism to account for the effects on a segment (S_n) which might be produced by the effort required for production of the following segment (S_{n+1}), and a "compatibility" mechanism which could adjust the amount of activity for a particular segment (S_n) depending upon the activity of the muscles in the preceding (S_{n-1}) segment.

The primary effect of the anticipatory mechanism would be to increase the activity for S_{n-1} in anticipation of a large amount of activity for segment S_n. The compatibility mechanism, on the other hand, would tend to reduce the activity for a segment following one with a large amount of activity.

The data of this study provides support for the existence of such mechanisms. In the posterior genioglossus muscle, for example, the largest amounts of activity were produced by the high vowels /i/ and /u/ . Table 3.2 indicates that the /k/ before these vowels evidenced greater activity than the /k/ before /ɑ/ , where no activity of the posterior genioglossus was present. Table 3.4 indicates that the correlation between the vowel activity and the consonant activity in this case was .60, indicating that an anticipatory mechanism might be involved. On the other hand, the correlations between the vowel (/i/ and /u/) peaks and the following consonant peaks, and between the second consonant and /g/ peaks were all negative, which is in accord with the predictions of the compatibility mechanism. It appears that the activity for the (stressed) vowel in these utterances tends to determine the amount of activity of the posterior genioglossus for the surrounding consonants. Of particular interest in these data is the fact that for the posterior genioglossus muscle, the presence of activity for the vowels tended to influence the level of consonant activity. The reverse effect, namely the consonants influencing the vowel articulation, was not found. In Table 3.2, the figures in parentheses in column 3 represent the averaged activity for these vowels (/i/ or /u/) in the labial environment /p___p/ . Although there is some fluctuation in the vowel emg amplitudes, it is rather small and is not systematic. These anticipatory and compatibility mechanisms might provide a physiological basis for the phenomena of regressive and progressive assimilation. This will be discussed in more detail in Chapter 4.

The behavior of the mylohyoid muscle cannot, however, be easily explained in terms of such anticipatory and compatibility mechanisms. The mylohyoid is minimally active for vowels, so the consonant peaks in the activity of this muscle cannot be adjusted depending upon the amount of activity in immediately preceding or following segments. The amount of mylohyoid contraction may be regulated by the activity of other muscles, however. Its activity is quite different from that of the posterior genioglossus in that it is more active for consonants in the immediate environment of back vowels, while the posterior genioglossus is most active for consonants in the environment of (preceding) high vowels. The mylohyoid activity does correspond somewhat to the styloglossus activity, however. Both muscles appear to be most active for consonants in the environment of back vowels. It is possible that there is some degree of synergy between the mylohyoid and styloglossus. Cortico-cerebellar interaction would probably be involved in the regulation of synergistic groups such as this.

The behavior of the mylohyoid muscle is much less precise than that of the posterior genioglossus. This can be seen by comparing the variability of the peak averaged emg amplitude of these muscles in Tables 3.1 and 3.2. This is intuitively understandable in terms of the function of these two muscles. The action of the posterior genioglossus serves to "bunch" the tongue. In doing so, however, it places the hump of the tongue under various places on the hard and soft palate, depending upon the amount of activity. Greater activity of the posterior genioglossus will lead to a more fronted articulation of a velar stop, less activity will result in normal or possibly retracted articulation. Since these articulations will have acoustic consequences, the activity of the genioglossus must be controlled fairly precisely. The mylohyoid is not under such an acoustic constraint, however, since its main function is simply to provide closure of the tongue against the roof of the mouth. Thus targets for the mylohyoid can be specified in a much looser manner than the targets for many of the other muscles. This accounts for the greater variability of the mylohyoid activity in acoustically and articulatorily similar gestures (/ts/ and /g/ in the frame, for example), and is an excellent illustration of the economy so often found in the motor control of speech.

Chapter 4

MODELS OF SPEECH PRODUCTION

In the previous chapter it was suggested that the motor commands to the muscles of the tongue are highly specific at the time that they are issued from the motor cortex. Another way of stating this is to say that different specifications for the muscular activity are issued from the motor cortex for different "intrinsic" allophones of the same phoneme. To return to a by-now familiar example, the posterior genioglossus muscle receives a different control signal for the fronted allophone ([k˂]) of the velar stop than it does for the normal or retracted allophone ([k]) or [k˃]). It was also suggested, following MacNeilage & DeClerk (1969), that these control signals may be different as a function of certain modifications, made just before the signals are issued to the muscles. These modifications may be made by a look-ahead procedure (the anticipatory mechanism), or by a mechanism which adjusts the "amplitude" of the motor command to make it compatible with the current state of the muscular system, using feedback to the motor areas from the cerebellum, which is "in touch" with the muscles *via* well-developed sensory feedback channels. The role of peripheral feedback (using monosynaptic spindle-afferent to motor neuron loops or a cerebellar loop) in speech production was minimized. It was also suggested that the gamma-efferent system (in the "gamma-drive" mode) does not play an important role in setting and regulating speech motor activity on a segment-by-segment basis, although it must be conceded that the evidence that this is so is, at present, not very strong. However such evidence as exists in favor of the gamma efferent system as an important mechanism in speech production is rather weaker.

As far as general modelling strategies go, there are two approaches which seem to be taken most commonly by speech production theorists. One approach has emphasized highly specific, quantitative models of vocal tract behavior in speech. Articulatory models of this sort have been developed by Stevens and House (1955), Henke (1966), Öhman (1967a), and Houde (1967). While these models frequently offer good insights into the articulatory behavior of the vocal tract, they are usually very non-specific about the underlying control mechanisms involved, and about the relation between the input and linguistic elements. Many (if not most) speech production models of this sort have claimed that the input to the model is a string of phonemes. Another approach has involved the construction of non-quantitative models of a much more abstract nature. Such models have been proposed by Kozhevnikov and Chistovich (1965), Cooper (1966), Öhman, et al. (1967), Fromkin (1968), Tatham (1969), Wickelgren (1969), and MacNeilage (1970). These "abstract" models are usually much more specific about the relation between units and processes of speech production and linguistic units. Models of both types are seldom proposed in a way that makes the neurophysiological correlates clear, although possible neurophysiological correlates of various parts of the models have been discussed by Kozhevnikov and Chistovich, Fromkin, Tatham, and MacNeilage.

One of the greatest defects of the less abstract quantitative models is that they tend to have few if any specific implications for linguistically or neurophysiologically oriented research in speech production. Henke (1966), for example, rejects an earlier model of tongue shape proposed by Stevens and House (1955) on the grounds that it is essentially a geometric artifice, which attempts to derive tongue shape on the basis of parameters and functions which have no conceptual significance. While Henke does indicate some aspects of his model which might have neurophysiological implications, he has not made a systematic attempt to correlate these aspects of his model with other models of a more abstract sort.

Öhman (1967a) proposed a quantitative model to account for the coarticulation of apical and dorsal stops with the extreme vowels /i/, /a/, and /u/. His model involved the postulation of target shapes for the tongue for vowels and consonants, and the determination of a coarticulation function operating over time which would predict the deviation from ideal consonant target shape in a VCV sequence as a function of the vowel targets. The model achieved success for predicting the total tongue shape for apical stops. However the dorsal stop (/g/) was found to require at least two target positions, depending upon whether the vowels in the environment of the stop were back (/u) or front (/i/). Ohman notes that the model suffers from the "attempt to represent the various lingual articulators *directly* in terms of parameters of [vocal tract] shapes." (p. 318) That is, the model could have been more insightful if the effects of motor commands upon the articulatory subsystems had been incorporated, rather than dealing with unified vocal tract shapes.

It seems clear that future work on speech production models must take into account the *motor* behavior of the vocal tract. Models which attempt simply to generate vocal tract shapes in the mid-saggital plane are not constrained enough to reveal many interesting facts about the act of speaking.

In the modelling of speech production, there have been at least three ways of viewing the abstract conceptions which underlie the observable patterns of muscular activity in speech. One way, which could be called the *direct motor command* model, assumes that there is associated with each abstract segment of speech (usually, and perhaps mistakenly, called a "phoneme") a set of basic motor commands for the muscles. For example, part of the set of motor commands for the abstract segment /p/ involves activation of orbicularis oris muscle to close the lips, activation of certain muscles involved in closing the jaw (such as the digastric), inhibition of muscles antagonistic to lip closure (such as the levator and depressor muscles of the upper and lower lips, respectively), activation of the muscles opening the glottis, etc. Articulators such as the tongue, which are not directly involved in the gesture, do not receive direct activating or inhibitory commands, but are usually described as being "free to coarticulate."

The "neuro-muscular invariance hypothesis," discussed in Chapter 3, can be viewed as a simple form of this type of model. However more sophisticated (and realistic) versions of the direct motor command model recognize that the basic motor commands can be modified by feedback, accounting for the variability found both in articulatory shapes and in the muscle contraction patterns revealed by EMG studies.

Another possible way of viewing the speaker's muscular activity in speech is to claim that he is trying to achieve certain target vocal tract shapes. In this model, which will be called the *target command* model, there is assumed to be a distinct spatial configuration of the articulator (or articulators) which exists as an ideal goal for the speaker for a particular segment, as a function of both the targets involved and of the current and anticipated state of the muscular system.

MacNeilage (1970) has reviewed the arguments in favor of the target command model (as opposed to the direct motor command model). The major points of MacNeilage's paper can be summarized as follows:

(1) Phonemes provide a good basis for models of speech production. That is, they provide a possible input to a speech production model.

(2) The motor commands associated with some phonemes are not invariant. That is, production of the phoneme in different phonetic contexts will involve different muscle gestures.

(3) Serial ordering of speech is effected at this level, that is, the level of phoneme targets.

(4) The targets are framed in terms of a space coordinate system. Such a system is claimed to have a reasonable degree of neurophysiological plausibility.

(5) A model is constructed to represent the conversion of phonemes into muscle movements. The essential properties of this model are illustrated by the flowchart of Figure 4.1. (The flowchart in this figure is my own graphical interpretation of MacNeilage's model.)

This model is claimed to be primarily open-loop, at least down to the level of the Motor System Control Mechanisms. These are context-invariant, and are modified by feedback from the muscles and other receptors.

In this model too, the gamma efferent system, particularly in what we have called the "gamma-drive" mode, plays an important role. It is considered to be one of the basic mechanisms by which the invariant movement command patterns can produce variable muscle actions, although it is by no means claimed to be the only one.

Another possible way of looking at the speech production process is to consider that the speaker's intention is to produce a particular acoustic quality for a segment, and that in doing this he is free to use his muscles in any way he can, just so long as he produces an auditorily acceptable sound. It is probable that the child, as he is learning to speak, is doing just this at one stage in his development, however it seems unlikely that the normal adult speaker controls his vocal tract in this manner. This would represent an extremely inefficient use of the marvelous complexities of the motor system.

A somewhat extreme version of the direct motor command model has been proposed by Wickelgren (1969). Wickelgren proposes that the motor behavior of speech is governed by what he calls "context sensitive elementary motor responses (*emr's*)." A context sensitive *emr* is essentially equivalent to the concept of intrinsic allophone discussed above, with a few small differences that will be apparent shortly. Thus the motor behavior of the vocal tract for the word *cat* would be determined by the ordered set of context sensitive *emr's* / ${}_{\#}\mathrm{k}_{æ}$, ${}_{\mathrm{k}}æ_{\mathrm{t}}$, ${}_{æ}\mathrm{t}_{\#}$ /. Given this hypothesis, we would expect that the motor behavior for different allophones of the same abstract phoneme would be different. In fact, Wickelgren claims that "the 'same phoneme' in different phonemic environments can be and must be different in some respects at all levels of the speech process, including the acoustic, vocal tract, and motor neuron levels." (p. 11) Although a considerable degree of variance depending upon context is usually found at the level of motor behavior, it is not the case that the behavior of the muscles is always different for different "context sensitive *emr's.*"

In EMG experiment #1 (see Chapter 2) two of the *S's* were asked, at the end of the experiment, to produce 20 repetitions of each of the near homophonous phrases "saw caught" and "Salk ought." The phrases were not pronounced

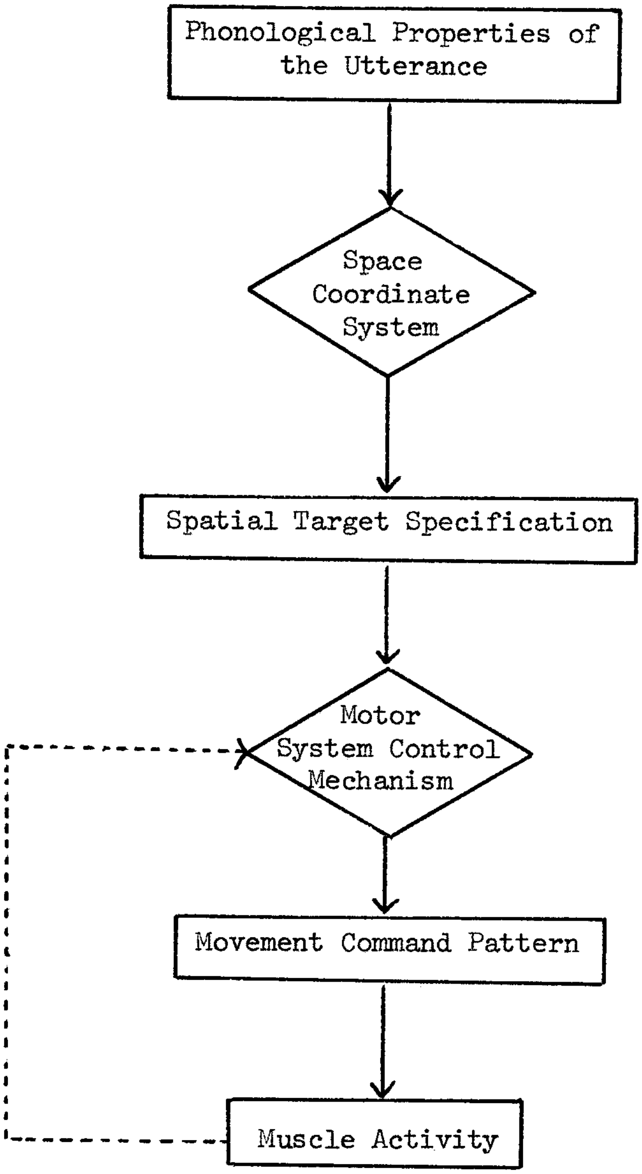


Figure 4.1

Graphical interpretation of the speech production model proposed by MacNeilage (1970). The dotted line represents sensory feedback paths.

consecutively, as were the sentences containing the other (CVC) items, but were mixed up together. In a traditional phonemic representation, these phrases differ only in that the word boundary (or juncture) /#/ is located in different positions: /sɔ#kɔt/ and /sɔk#ɔt/. Acoustically, they differ very little. In "Salk ought" the first vowel is slightly shorter than in "saw caught," and the velar stop closure is slightly longer. In Wickelgren's terms these utterances would presumably be governed by the *emr* sequences /${}_{\#}s_{ɔ}$, ${}_{s}ɔ_{\#}$, ${}_{\#}k_{ɔ}$, ${}_{k}ɔ_{t}$, ${}_{ɔ}t_{\#}$ / ("saw caught") and /${}_{\#}s_{ɔ}$, ${}_{s}ɔ_{k}$, ${}_{ɔ}k_{\#}$, ${}_{\#}ɔ_{t}$, ${}_{ɔ}t_{\#}$/ ("Salk ought"). The two velar stops in these phrases have different representations in terms of *emr's*, although they are clearly the same phoneme. Figure 4.2 shows the averaged EMG activity for the posterior genioglossus and mylohyoid muscles, as well as the averaged audio signals, for one of the *S's* pronouncing these two phrases. Data for the other *S* is very similar. The posterior genioglossus and the mylohyoid (along with the styloglossus) are the muscles most important in producing velar stops. As is clearly shown in the figure, the only significant difference between these EMG signals is in the timing. The average amplitudes for both instances of the /k/ for both muscles are identical. The timing differences between these two utterances are a reflection of the difference in syllable structure. Vowels in open syllables are longer, *ceteris paribus,* than vowels in closed syllables. Of course it is possible that some of the other extrinsic tongue muscles involved in the production of velar stop closure (such as the styloglossus) or in velar stop release (presumably the hyoglossus) would show systematic differences in amplitude for these two instances of the phoneme /k/, however this seems unlikely.

Since Wickelgren claims that "...context sensitive coding is essentially equivalent to coding in terms of overlapping syllables." (p. 7), it is difficult to understand how his model could predict the timing differences in the "emr's" in these two phrases. In any event, his claim that different context-sensitive *emr's* must have different realizations at the motor level is not confirmed by these data from the tongue muscles.

There are other problems with Wickelgren's model, most of them revolving around his claim that the serial ordering of motor behavior is effected at the level of his context-sensitive *emr's*. Some of these problems have been discussed by MacNeilage (1970), Whitaker (1970), and Tatham (1970). In summary, the major problem with this model is that it is simply not abstract enough to allow significant generalizations to be made. An interesting model of speech production will have to make a number of generalizations similar to those made, usually at a more abstract level, by a set of phonological rules. No phonologist would consider the statement that vowels in English are lengthened before d, and b, and g, and m, n, l, etc. very interesting. The phonologist must abstract out the *feature* common to all those segments and state the facts as a general rule. It seems certain that a

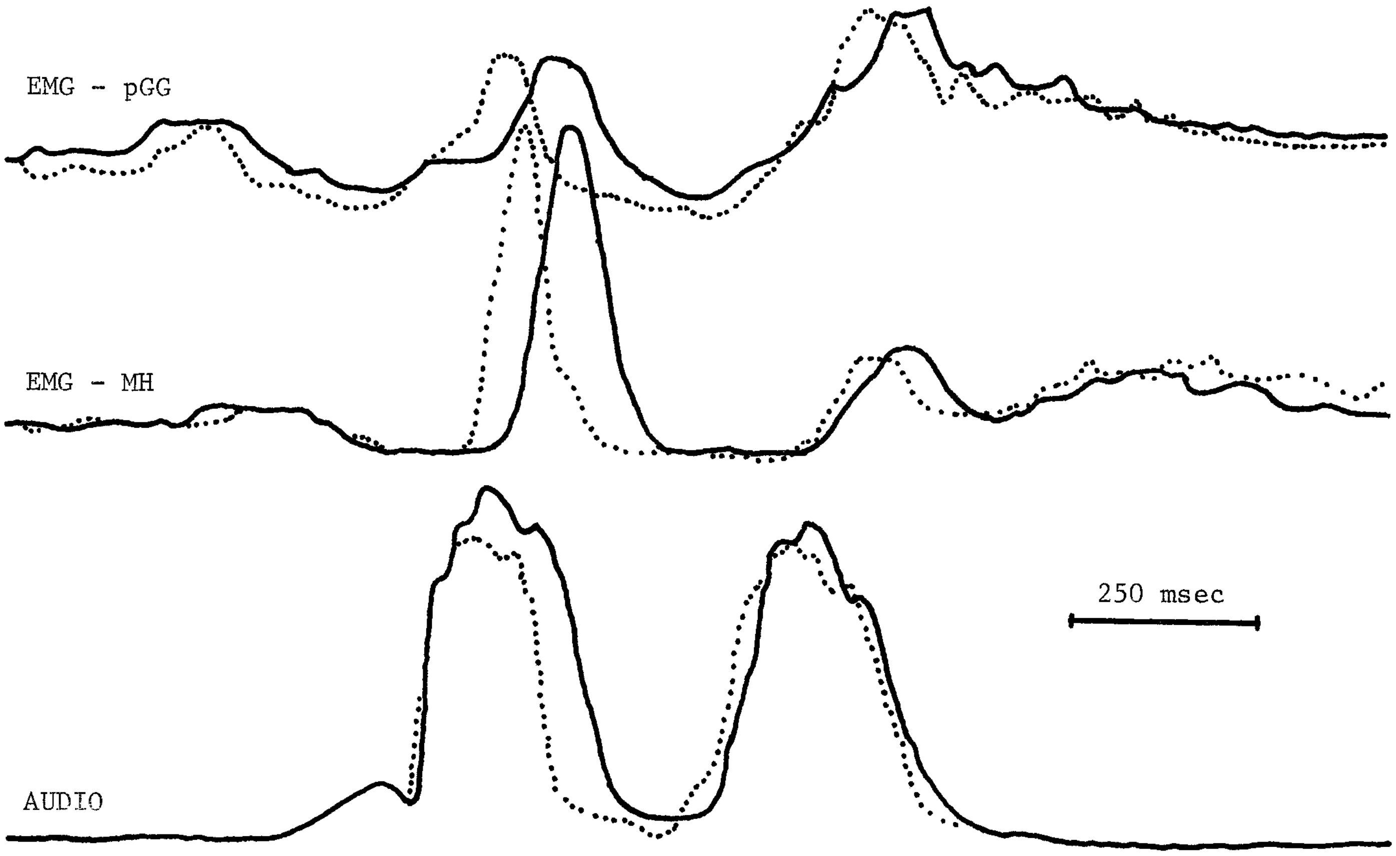


Figure 4.2

Averaged EMG signals (n=20) from the posterior genioglossus and mylohyoid muscles for the utterances "saw caught" (solid lines) and "Salk ought" (dotted lines). Averaged audio signals are shown below the EMG curves.

correct model of speech production will also have to incorporate rule-like statements in describing the interdependencies (contextual effects) among segments (and even features of segments). Wickelgren's model essentially prevents such generalizations from being made.

There are, then, at least four possible varieties of the *direct motor command* model, each making different empirical claims. The first assumes context-free (phonemic) coding at the motor level, with variability being a result of coarticulation at the articulatory and/or acoustic levels. This is the model referred to above as the neuromuscular invariance hypothesis. The second variety, represented by Wickelgren's model, assumes context-sensitive (allophonic) coding at the motor level. The third variety would assume invariant phonemic coding at the level of the motor cortex, but would recognize that the actual motor commands to the muscles differ for the same phoneme in different phonetic contexts. However this variant would ascribe the differences in motor commands to the action of peripheral feedback mechanisms. The gamma-efferent system normally plays an important role in this variety of the direct motor command model as it does in the target command model. A fourth variety of the direct motor command model would assume that at some very abstract level there exists "idealized" motor commands, which are identical for each instance of a phoneme. This model would claim that the idealized motor commands can be modified before they are actually issued from the motor cortex. This variety might claim that peripheral feedback can also be employed to modify the motor commands.

Of course, it must be realized that a model of speech production, whether it involves the spatial targets or identical motor commands as the abstract invariant units, could postulate that both peripheral feedback and cortical reorganization are involved in either realizing the "surface" motor commands that actualize the target (target command model) or in modifying idealized invariant motor commands.

It has already been pointed out that the first three of these versions of the direct motor command model are inadequate in various ways. The fourth version is, in a sense, the weakest version, since it makes few if any empirical claims. Given the current state of the art in experimental phonetics, there does not appear to be any direct way of distinguishing empirically between the claims that abstract segments (or features) are stored as idealized vocal tract shapes (target command model) or that they are stored as sets of idealized motor commands. Experiments using peripheral data such as EMG, X-ray films of articulatory dynamics, and acoustic events do not seem to be powerful enough to allow us to distinguish between the two alternative models. Furthermore, it is difficult to envision new experimental techniques which could be developed to test these alternatives. The level at which speech events are straightforwardly correlatable with abstract linguistic units seems to have been pushed beyond the reach of currently available techniques, and may have been pushed beyond the reach of any technique that can be used with live human subjects.

LINGUISTIC UNITS AND MOTOR BEHAVIOR

In the previous sections the question of the abstract representation of the units governing the motor behavior of the VT was pursued without involving any serious specific discussion of the relation of these abstract units to units of linguistic representation. The terms "phonemes" and "allophone" were used, albeit quite loosely. We shall not turn to a more detailed consideration of the nature of the relationship between linguistic units and processes (i.e., rules) and the data and processes of a speech production model. These considerations also hold the promise of shedding some light on the question posed above, that is, is the motor behavior of the vocal tract governed by spatial target specifications which are converted into motor commands or by idealized motor commands directly.

Phonologists have treated the question of the relation between linguistic units such as features and vocal tract behavior in a somewhat vague manner. Stevens and Halle (1967) consider that "the segments and features of the abstract representation as instructions for particular types of behavior of the speech-generating mechanism." (p. 90) Postal (1968) claims that "phonetic representation is . . . part of a theory about the instructions sent from the central nervous system to the speech apparatus . . ." (p. 6). Further on, he writes that "such linguistic equipment [feature matrices] is descriptive of the ideal behavior of the vocal apparatus with respect to its production of speech." (p. 59). Chomsky and Halle (1968) regard the phonetic features as" . . . physical scales describing independently controllable aspects of the speech event, such as vocalicness, nasality, voicing, glottalization. There are, therefore, as many phonetic features as there are aspects under partially independent control." (p. 297) All this does not help much. About all that can be gleaned from a survey of the recent literature in linguistics is that abstract segments are composed of features which are essentially mentalistic, but which have real physical correlates in that they represent instructions to the vocal apparatus. If one examines the actual features proposed (e.g. in Chomsky & Halle 1968) one finds that these claims regarding the status of the features are often either self-contradictory or totally empty of any empirical content. For example, the feature [stress] is considered to be an abstract feature, the value of which is often largely determined by the syntactic shape of the string being analyzed. Some specifications of this feature may not have any physical (physiological or acoustic) reality, and thus can hardly be viewed as "describing an independently controllable aspect of the speech event," where "speech event" has some physical interpretation. Or consider the feature [anterior). As Fant (1969) noted, this feature does not, by itself, define a shape or position of the articulators. [anterior] refers to articulations (i. e., obstructions) which are produced anterior to an arbitrary dividing line somewhere behind the alveolar ridge. Labial obstruents are specified as [+anterior, -coronal], and are distinguished from the [-coronal] velars and uvulars by this feature. It seems implausible to consider the tip of the tongue and the lips as one articulatory sub-system, yet this is what such a feature forces one to do, if the features are to have any neuro-muscular implications. Also, consider the impossibility of treating labial-alveolar compound stops ([p͡t]), such as are found in Margi (Ladefoged 1964b), with such features as [anterior] and [coronal]. Despite these objections, there are a number of linguistic features which would appear to have interesting neuromuscular implications.

Most of the features described in Chomsky and Halle (1968) (henceforth SPE) are articulatory features, that is, they define configurations of the articulators in the vocal tract. Even the major class features, such as [sonorant], [vocalic], and [consonantal], are defined in terms of vocal tract shapes. Two conventions are needed to interpret the articulatory correlates of the SPE features: *spontaneous voicing* and the *neutral position of the vocal tract*. We will not be concerned with the first, and somewhat speculative convention in this study. It is used to define the feature [sonorant]. The neutral position of the vocal tract in SPE is considered to be the position adopted by a speaker just prior to beginning an utterance (cf. Perkell 1969). In English, the neutral position is considered to be the tongue position of the English vowel [ɛ] in *bed*. Thus mid front vowels are neutral, and all deviation from this neutral position is handled by positive or integer specification of one of the tongue body features: [high], [low], or [back]. The physiological analog of this articulatory neutral position could be considered to be the basic tonus which many muscles adopt just before speech (Öhman 1967b).

The features with which we will be mainly concerned are [consonantal], [vocalic], [coronal], [high], [low], and [back]. The correlates for these features as given in SPE are listed below:

[consonantal] - consonantal sounds are produced with a radical obstruction in the midsaggital region of the vocal tract.

[vocalic] - vocalic sounds are produced with an *oral* cavity in which the most radical constriction does not exceed that found in the high vowels [i] and [u] and with vocal cords that are positioned so as to allow spontaneous voicing; in producing nonvocalic sounds, one or both of these conditions are not satisfied.

[coronal] - coronal sounds are produced with the blade of the tongue raised from its neutral position; noncoronal sounds are produced with the blade of the tongue in the neutral position.

[high] - high sounds are produced by raising the body of the tongue above the level that it occupies in the neutral position; nonhigh sounds are produced without such a raising of the tongue body.

[low] - low sounds are produced by lowering the body of the tongue below the level that it occupies in the neutral position; nonlow sounds are produced without such a lowering of the body of the tongue.

[back] - back sounds are produced by retracting the body of the tongue from the neutral position; nonback sounds are produced without such a retraction from the neutral position.

(SPE, pp. 302-305)

The features [coronal], [high], [low], and [back] suffice to describe the principal articulatory positions of the tongue. Other features, such as [anterior] and [distributed], handle the position and length of a closure or obstruction in the vocal tract.

Let us assume, as a specific modelling strategy, that the features described above represent spatial target positions of the vocal tract. Thus we could view the features (appropriately specified) as possible inputs to a *target command* model of speech production. What we then need to describe is the mechanism by which the target commands are translated into actual motor commands to the muscles. Ideally, we would like to do this in a fully quantitative manner. However, we do not yet understand enough about quantizing the EMG activity which represents the actual motor commands to the muscles, so our approach will be only semi-quantitative.

The set of processes which convert target commands (i.e., properly specified features) into motor commands we will call, following Laver (1968), the "myodynamic" rules. Figure 4.3 indicates the general plan of the model.

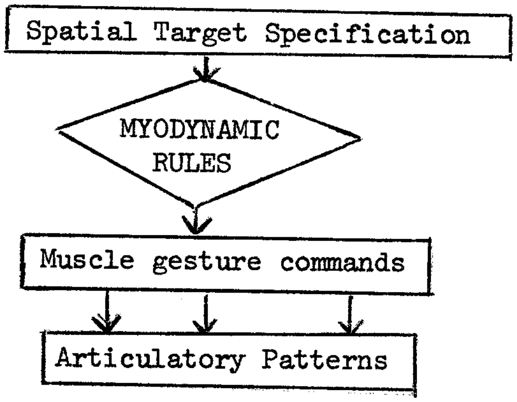


Figure 4.3

Since the myodynamic rules represent a conversion from abstract spatial targets to gesture commands, they must in effect assign values representing amount of contraction to the muscles involved in controlling the tongue. Thus we could consider the muscles themselves as the "features" at this level, and view the myodynamic rules as rules which are of the general form

$$(1) \qquad \begin{bmatrix} n\ F_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ n\ F_n \end{bmatrix} \longrightarrow \begin{bmatrix} m\ M_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ m\ M_n \end{bmatrix}$$

where $[F_i]$ is a spatial target feature, and n is a plus, a minus, or an integer value specification of this feature, and $[m\ M_i]$ represents a numerically specified amount (m) of contraction of a muscle M_i.

We will determine the values which a muscle M_i can take in the following way. The m of $[m\ M_i]$ can be specified to be any of the integral values 0 to 4, in the following general manner.

$0\ M_i$ - muscle M_i is inhibited.

$1\ M_i$ - muscle M_i is contracting weakly. In terms of the data reported in Chapter 2, this would consist of only a few motor units firing. This is found, for example, in the contraction of the mylohyoid muscle for vowels.

$$\left.\begin{array}{l} 2\ M_i \\ 3\ M_i \\ 4\ M_i \end{array}\right\}$$ Three values of contraction for events when the muscle is producing an interference pattern.

The choice of 4 levels of contraction was made merely by carefully inspecting the data reported in Chapter 2. There are no external criteria which could be used to motivate this choice. It seems that at least 4 levels are needed, to distinguish the contraction pattern of the mylohyoid muscle (MH), for example:

0 MH - low vowels (/a/)

1 MH - high vowels (/i/)

2 MH - [+ coronal] stops (/t/)

3 MH - fronted /k/ ([$\underset{<}{\text{k}}$])

4 MH - "normal" /k/ ([k])

Additional data could of course force a revision of this classification. However, it seems likely that at least four levels of contraction, plus inhibition, are required to handle the data of experiments 1 and 2 in Chapter 2. In any case, the number of levels which can be specified will probably not affect the statement of the myodynamic rules in any major way.

In Table 4.1, the data from four of the muscles studied in the previous experiments is summarized. This table indicates which segment types fall into certain contraction value classes for each muscle. The muscles (hence, gesture command features) are the posterior genioglossus (pGG), mylohyoid (MH), styloglossus (SG), and the glossopharyngeal (GP). In this table, values have been indicated for the following segments:

t_1	the [t] before a low vowel (/a/) of experiment #1, Chapter 2.
t_2	the [t] before a high front vowel (/i/) of experiment #1.
k_1	the normal velar stop [k] before a low vowel (experiment #1 and #2).
k_2	the fronted velar stop [k] before a high or mid front vowel (experiment #1 and #2).
w j	the approximants (glides) which follow the nonlow vowels of experiments #1 and #2 (/i e o u/).

The vowels are listed in this manner: vowels specified for MH and pGG activity are given the values which they attained in the labial environment /p__p/. We will plot activity only for the vowels /i/, /e/, /ɑ/, /ɔ/, /o/, /u/. For the SG and GP muscles, the situation is more complex. Recall that the S in experiment #2 pronounced the vowels /i a ɑ ɔ o u/ after /b/ and gave these vowels their cardinal values. Furthermore, he produced a number of vocalic nuclei in various utterances which were more typical of his normal speech. In order to make the results compatible with those for the MH and pGG muscles, we will consider only the nuclei [e^{j}] ("keg"), [i^{j}] ("peeve"). [ɑ] ("fob"), [ɔ] ("paw"), [o^{w}] ("cope"), and [u^{w}] ("poop"). Therefore, the entries under the vowel symbols of Table 4.1 represent data from two different experiments. The subject of experiment #2 (the SG and GP muscles) speaks a variety of British English, and his vowels are somewhat different auditorily from those produced by the American subjects of experiment #1. Nevertheless, they are basically similar in articulatory terms, the greatest difference probably being in the low and back vowels /a/ and /ɔ/. The vowel in "fob" [*fab*]) for the subject of experiment #2 is lower and more backed than the vowel in "pop" ([*pap*]) for the speaker of General American English. The vowels in "paw" (experiment #2) and /pɔp/ ("paup") (experiment #1) are not too important within the structure of this table, however, since low vowels do not employ the pGG (and MH) muscles investigated in experiment #1. One aspect of Table 4.1 which is highly subjective is the assignment of the amplitudes for the vowel part of the vocalic nuclei [i^{j}], [e^{j}], [o^{w}] and [u^{w}]. The peaks of activity for these nuclei for the pGG and SG muscles were typically caused by the glide portion of the nucleus. In order to systemize the assignment of vowel amplitudes, we decided to consider the amplitude of the averaged emg signal at the absolute peak to indicate the activity for the glide [j] or [w], and the activity about 75 msec. prior to the peak to represent the vowel. This procedure is rather arbitrary, and clearly should be improved. Nevertheless, the data will be at least indicative of the contraction for the vowel.

	t_1	t_2	k_1	k_2	j	w	i	e	ɑ	ɔ	o	u
MH	2	2	4	3	1	1	1	1	0	0	0	1
pGG	1	2	2	3	4	4	3	2	0	0	2	3
SG	-	-	4	3	0	3	0	0	0	0	2	2
GP	-	-	0	1	0	0	0	0	3	4	2	0

Table 4.1

Table 4.1 gives, then, the values which should be derived by the myodynamic rules. It is now necessary to determine the input values for these rules, in terms of specification of the relevant major-class and tongue body features. At the systematic phonemic level, these segments would have the plus and minus values indicated in Table 4.2. In English (and most other languages), the feature [coronal], for vowels and glides is redundantly [-coronal], since there are no retroflexed vowels or glides, and the feature [low] for true consonants is redundantly [- low], since there are no pharyngealized consonants. Nevertheless, these features must be specified at the phonetic level.

	t_1	t_2	k_1	k_2	j	w	i	e	ɑ	ɔ	o	u
cons	+	+	+	+	-	-	-	-	-	-	-	-
voc	-	-	-	-	-	-	+	+	+	+	+	+
cor	+	+	-	-	-	-	-	-	-	-	-	-
high	-	-	+	+	+	+	+	-	-	-	-	+
low	-	-	-	-	-	-	-	-	+	+	-	-
back	-	-	+	+	-	+	-	-	+	+	+	+

Table 4.2

There are a number of problems inherent in determining the exact specification for these segments at the systematic phonetic level. We shall first review the mechanisms by which segments at this level could receive integer values. Postal (1968, p. 65ff) suggests that the phonological component of a grammar contains a set of phonetic Detail Rules which map plus and minus values of the phonological/phonetic features into integer values. According to Postal, these rules have the general form indicated in (2) below:

(2)

$$\left\{\begin{array}{ccc} +\ F_i & \longrightarrow & 1\ F_i \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & \longrightarrow & \cdot \\ \cdot & & \cdot \\ -\ F_i & \longrightarrow & n\ F_i \end{array}\right\} \text{ in } X\_Y$$

where X_Y is a schema for the environments of the rule. Postal further suggests that there is probably no overlapping in detail rules, that is there is no n-ary phonetic value which is a possible mapping of both plus and minus values. The language specific detail rules are a sub-set of universal phonetic detail rules.

The environment schema of a phonetic detail rule can be considered as the expansion (3a), (3b), (3c)

(3a)

$$/\ \underline{\qquad}\ \begin{bmatrix} n\ F_i \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ n\ F_n \end{bmatrix}$$

(3b)

$$/\ \begin{bmatrix} n\ F_i \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ n\ F_n \end{bmatrix}\ \underline{\qquad}$$

(3c)

$$/\ \begin{bmatrix} \underline{\qquad\quad} \\ \\ n\ F_i \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ n\ F_n \end{bmatrix}$$

where *n* is a plus, a minus, or an integer value, and F_i is a phonological/ phonetic feature.

These environment schemata show that features can be assigned integer values depending upon values of the same or other features in adjacent segments, or depending upon the values of other features in the same segment. As an example of (3c), consider the detail rule which would be necessary to assign integer values to the feature [high] for various segments. [k], [w], and [u] are all [+ high] segments. In purely physical terms, however, the tongue is higher for an obstruent than for a glide or a vowel, and is higher for a glide than for a [+ high] vowel. If we decide that velar obstruents are [3 high], glides are [2 high], and high vowels (in English) are basically [1 high], then a rule of the following sort is needed:

(4)
$$[+\text{ high}] \longrightarrow \left\{ \begin{array}{l} [3\text{ high}] \;/\; \begin{bmatrix} \underline{\qquad} \\ -\text{ son} \end{bmatrix} \\[2em] [2\text{ high}] \;/\; \begin{bmatrix} \underline{\qquad} \\ -\text{ cons} \\ -\text{ voc} \end{bmatrix} \\[2em] [1\text{ high}] \;/\; \begin{bmatrix} \underline{\qquad} \\ +\text{ voc} \end{bmatrix} \end{array} \right\}$$

Other adjustments would have to be made to differentiate [+ continuant] and [- continuant] obstruents. This rule is considered to be language specific, not universal, to account for phonetic differences between languages, such as the fact that tense high vowels are higher (and more tense) in French than in English, or that in Korean glides are not necessarily higher than the high vowels (Lee, personal communication).

Other detail rules will handle phonetic facts which are dependent upon feature values in the adjacent segments (schemata 3a and 3b). Consider, for example, the treatment of velar fronting. In English, velar stops are fronted before high and mid front vowels. This fact could be handled not by a detail rule but by a phonological rule such as (5)

(5)
$$\begin{bmatrix} -\text{ continuant} \\ +\text{ high} \end{bmatrix} \longrightarrow [-\text{ back}] \;/\; \underline{\qquad} \begin{bmatrix} -\text{ cons} \\ -\text{ back} \\ -\text{ low} \end{bmatrix}$$

(cf. Bach 1968). This treatment is possible in English, since English has no true palatal consonants, i.e., consonants which are $\begin{bmatrix} +\text{ high} \\ -\text{ back} \end{bmatrix}$. Rule (5), however, claims that a /k/ becomes [c], a palatal stop, before high front vowels, which is not the correct phonetic statement (particularly since the phonological component of a grammar probably includes a convention which turns palatal stops into affricates). It would be more appropriate to claim that it is less backed before an /i/, not that it is nonback. Rule (5) would probably be better stated as a detail rule. If we assume that uvulars are basically [4 back], "normal" velars are [2 back], and "fronted" velars are [1 back] (these values will be explained below), then (5) could be stated as the detail rule

$$(6) \qquad \begin{bmatrix} -\ \text{continuant} \\ +\ \text{high} \end{bmatrix} \rightarrow [1\ \text{back}] \;/\; \underline{\qquad} \begin{bmatrix} -\ \text{cons} \\ -\ \text{back} \\ -\ \text{low} \end{bmatrix}$$

However, this rule does not seem to be capturing the correct phonetic facts either. The fronting of velars before vowels in English is a function of the backness of the vowel. We would desire a rule which a assigns the assimilated backness values to the stop on the basis of the integer values assigned to the following vowel or glide for this feature:

$$(7) \qquad \begin{bmatrix} -\ \text{continuant} \\ +\ \text{high} \end{bmatrix} \rightarrow [\alpha\ \text{back}] \qquad \underline{\qquad} \begin{bmatrix} -\ \text{cons} \\ \alpha\ \text{back} \end{bmatrix}$$

where α is a variable ranging over integer values. There are a number of problems with this rule, however. Since /i/ and /e/ are nonback vowels, we want to be able to say that they are [0 back]. If rule (7) makes [k] [0 back], it is in effect making the same claim as rule (5).

Ladefoged (1967b) has emphasized that phonetic facts about the pronunciation of a language must be indicated at the systematic phonetic level. A complete grammar of Italian will reflect the fact that /t/, /d/, and /n/ in Italian are dental, rather than alveolar or retroflexed, whereas a grammar of English must state that English [+ coronal] stops are alveolar. McCawley (1968) also notes that the systematic phonetic level should reflect facts about utterances which produce a "native accent." Dental stops are part of a native accent in Italian, French, Spanish, or New York English, but not in General American.

The integer specification of articulatory features at the systematic phonetic level has not been thoroughly investigated. It is a topic that needs systematic investigation. In SPE, the only feature which receives integer specification at this level is [stress], by an extremely complex set of cyclically ordered rules. But it was implied above that[stress] is a very abstract feature. It has no simple articulatory or acoustic correlates. Lieberman (1965) has shown that trained linguists cannot recognize more than two degrees of stress in utterances in which the amplitude, timing, and fundamental frequency cues are present, but from which the syntactic and semantic information has been removed. The multitude of stress levels predicted by the rules of SPE may have some psychological reality (Ladefoged and Fromkin 1968), although this matter is still open to debate (Ohala 1970). Nevertheless, it is clear that the distinction between [3 stress] and [4 stress] is minimal, to say the least, at the level of articulatory gestures.

Few, if any, rules have ever been written which specify how place or articulation features such as those we are discussing here are given integer specification at the systematic phonetic level. For some of these features, such as [high], [back], and [low], it is possible to assign integer specifications reasonably easily. For others, it is very difficult to imagine how integer specifications could be assigned. The feature [coronal], for example, is defined in such a manner that it would be difficult to assign integer values to it in any reasonable way. It might appear that [+ continuant] coronal sounds like [s] would have a smaller integer value for this feature than coronal stops. But notice here that the "continuancy" or fricative quality of [s] (or [ʃ]) is a function of the groove in the center part of the tongue. The sides of the tongue are in a position very similar to that for [t] or [n].

It may be the case that the difference between [t̪] and [t] would be reflected by integer specification of the feature [coronal], however this contrast (which is found phonologically in e.g., Malayalam), is handled by the feature [distributed], in the SPE feature system, [t̪] being [- distributed], and [t] [+ distributed]. The feature [coronal] does appear to be correlatable with a gesture raising the blade of the tongue. This basic gesture, however, probably differs very little among different [+ coronal] consonants. As suggested by Strong (1956), other intrinsic muscles are called into play to perform the fine adjustments which distinguish sounds like [t] and [s].

	k_1	k_2	j	w	i	e	ɑ	ɔ	o	u
high	3	3	2	2	1	0	0	0	0	1
low	0	0	0	0	0	0	2	1	0	0
back	2	1	0	3	0	0	2	3	3	3

Table 4.3

Integer values for the tongue body features at the systematic phonetic level.

In Table 4.3, the segments which are distinguished by the tongue body features are given integer specification. We assume that the feature [back] takes 5 values, as shown in Table 4.4.

Back:	1	2	3	4	5
	k̠	k	ɔ	q	ʕ
		ɑ	o		
			u		
			w		

Table 4.4

The segments [q] and [ʕ] have not been investigated (in normal speech) in this study, so we will not discuss them further.

The values for [high] have been assigned essentially as described earlier (see Rule 4). For the feature [low], a distinction is made only between [ɑ] ([2 low]) and [ɔ] ([1 low]). The convention that $[-\,F_i] \rightarrow [0\,F_i]$ for each of the tongue body features has been adopted.

If the model schematized in Figure 4.3 is correct, then the matrix in Table 4.3 should serve as the correct input to the myodynamic rules, deriving the data in Table 4.1 as the output. It would be possible to assume that myodynamic rules for these features have the general form

$$[n\,F_i] \rightarrow \begin{bmatrix} m\,M_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ m\,M_n \end{bmatrix} \Bigg/ \;\; X\underline{\quad}Y \tag{8}$$

Condition: $n > 0$ (cf. (1) above)

This format states that a feature F_i having some integer value greater than zero is assigned contraction values (*m*) for muscles 1 to n in the environment X__Y. A set of rules of this form which maps specifications of Table 4.3 into the data of Table 4.1 would contain many ad hoc statements. But there are also some interesting generalizations that can be made on the basis of the data in these tables. These are summarized below:

1. The styloglossus (SG) muscle is active for segments which are non-low and back. The amount of SG activity is proportional to the values for [high]. For segments which are [3 high], SG activity is inversely proportioned to the value for the feature [back]. The SG activity for the vowel /o/ provides the only real exception to these statements.

2. Activity for the posterior genioglossus (pGG) muscle is found for all high segments, as well as the mid vowels /e/ and /o/. (There is also a small amount of pGG activity for the coronal stops t_1 and t_2.) pGG activity is not directly proportional to either the height or the backness of the segment. In general, the pattern seems to be that the pGG contracts strongly for [+ high] segments. The amount of activity is reduced if the segment is a mid vowel /e/ or /o/, or if the SG or mylohyoid (MH) muscles are markedly active. From a physiological point of view, this is easily understandable. Less activity of the pGG is required to make a velar stop than to make a high glide or vowel, since the MH muscle is also acting in the case of the stops. The expected absence of SG activity for the front mid vowel /e/ may also explain why the MH contracts slightly for this vowel, but not for /o/. If a velar stop is fronted (k_2) there is greater pGG activity and a reduction of SG activity.

3. The glossopharyngeal (GP) muscle is active for low and mid back vowels. Since [ɔ] is more backed than [ɑ], this muscle (which pulls the root or posterior third of the tongue towards the posterior pharyngeal wall) contracts more strongly for [ɔ] than for [ɑ]. [o] is also backed, but it is probably slightly higher than [ɔ]. The small amount of activity in the GP muscle for the fronted velar stop (k_2) is paradoxical. This might represent a stabilizing gesture antagonistic to the pull of the pGG muscle.

4. The behavior of the MH muscle does not appear to be governed by any of the tongue body features discussed here. It is markedly active only for obstruents. It is probable that the MH contraction is governed by the degree of mandible opening. The MH is less active for the fronted velar (k_2) than for the more neutral k_1 segments. In a sequence like /iki/ the mandible opens less than in a sequence like /aka/. (Amerman et al. 1970). If this hypothesis regarding the behavior of the MH muscle is correct, it suggests that the degree of mandibular opening, possibly computed as a function of the vocalic context (Ohala 1970), is a target relevant to the specification of the activity of the MH muscle. Information of this sort is not directly contained in the feature specifications at the systematic phonetic level.

So far in this discussion, it has been assumed that the transformation from the spatial target level (which we equated with the systematic phonetic level of a phonological description) to the muscle gesture commands could be stated by a set of discrete rules. General schemata for such rules were suggested in (1) and (8) above. Rules of this sort would achieve some success in stating the transformations involved.

However, notice that in assuming that the myodynamic rules operate on fully specified systematic phonetic representations, we are essentially restricting the myodynamic rules to be of the form

(9) $$[n\ F_i] \rightarrow [m\ M_i]\ /\ \begin{bmatrix} \underline{\qquad} \\ X \end{bmatrix}$$

where the within-segment environment can be features and/or muscle motor commands. That is, all myodynamic rules will assign muscle contraction values for features within one segment. However there seems to be little reason, *a priori*, to restrict the myodynamic rules in this way. If we assume that they can have the form

(10) $$[n\ F_i] \rightarrow [m\ M_i]\ /\ X\underline{\quad}Y$$

Where X__Y can be expanded as indicated in 3 a-c above, then the myodynamic rules can be used to state at least some of the phonetic between-segment context effects normally handled by the phonetic detail rules. That is to say, the spatial targets could be considered to be more abstract, and the element of the speech production model known as the myodynamic rules could account for most of the effects which are normally described by phonetic detail rules. If we assume that the myodynamic rules represent transformations that take place at a cortical level, and are not simply a reflection of effects that take place at a more peripheral level, then there does not seem to be any strong reason for constraining the myodynamic rules to operate in a restricted within-segment manner. That is, the myodynamic rules could operate on strings of elements, perhaps having hierarchical structure. The foregoing comments would hold, however, only if the assumption is made that the myodynamic rules are language-specific rather than universal. If the speech production model is considered to be universal, then the myodynamic rules would in all probability have to be constrained to operate only on systematic phonetic segments, rather than on strings, since context effects do seem to vary from language to language (although this is by no means well-established).

It is, in general, possible to consider that at least some of the phonological/phonetic features represent spatial targets, and that segments composed of features can be realized as motor commands on a feature-by-feature basis by myodynamic rules. It is not possible, however, to implement a model in which each feature has associated with it an idealized set of motor commands. This is because a feature having a certain specification does not by itself have any neuromuscular implications. Some of the features which govern the position of the body of the tongue do seem to have a fairly direct relationship to muscle activity. For example, all [+ high] segments appear to involve contraction of the posterior genioglossus muscle. The activity of other muscles, however, can be accounted for only with reference to combinations of features. The feature specification [+ back] is not associated directly with any specific muscle. Non-low, back segments involve activity of the styloglossus muscle, whereas low, back segments require inhibition of the styloglossus, and activation of other muscles, such as the glossopharyngeal. Phonological/phonetic features can be shown to be psychologically real in a number of senses. Numerous psychological studies have shown that features can predict errors in perception and confusion in short-term memory (Wickelgren 1965). It is reasonable to assume that features are directly involved in speech production also, although the features involved may not be exactly those proposed in SPE. The assumption that speech is encoded, at an abstract level, in terms of features leads one to prefer a *target command* model of speech production over a *direct motor command* model.

NATURAL RULES IN PHONOLOGY

It was mentioned above that the processes described in a set of phonological rules and the processes inherent in a speech production model could be partially parallel, and may overlap to some extent. This overlap seems to occur in the area of certain "phonetically natural" phonological rules. Among the phonological rules which have been considered by linguists to be "natural rules", the assimilative rules such as velar fronting, vowel nasalization, and consonant cluster place of articulation assimilation are predominant (Schachter 1969, Schane 1969, Bach and Harms 1969). Schane has noted that natural assimilative rules are those in which the primary features of classes of segments are assimilated. For vowels the primary features are [high], [back], [low], and [round]. For consonants, the primary features are ones such as [coronal], [anterior], [distributed], and [nasal]. Thus consonants tend to take on the tongue body characteristics of surrounding vowels, whereas vowels assimilate to the primary consonant features. The former process is evident in the fronting of velar stops, and the latter in the nasalization of vowels before (and occasionally after, see Schachter and Fromkin 1968) nasal consonants. These natural rules could also be reflected in speech production models. Consider the case of velar fronting. In the data of this study, it was noted that the activity of the posterior genioglossus muscle is increased for a velar stop before a high front vowel, and the activity of the styloglossus muscle is somewhat less in a similar environment. The articulatory consequence of this change in muscle activity patterns is to front the articulation of the stop.

At first sight, it may appear that this activity is a consequence of anticipatory processes of the sort discussed in Chapter 3. A large amount of activity for the vowel, such as that found in the posterior genioglossus for /i/, is anticipated in the articulation of the velar stop, having the articulatory consequence of fronting. These anticipations cannot be a function of the activity of one muscle only, however. The vowels /i/ and /u/ have approximately the same amount of activity for the posterior genioglossus muscle, but the activity in the posterior genioglossus is increased only for the /k/ before /i/, and not for the /k/ before /u/. Also, the reduction of the styloglossus muscle activity for a velar stop before a front vowel cannot be an anticipatory phenomenon. It also cannot be handled by the compatibility mechanism discussed in Chapter 3, since we would expect that the activity would be increased before a vowel having little or no styloglossus muscle activity if the compatibility mechanism were operating.

Consider another case of a natural phonological rule: vowel nasalization. Fritzell (1969) has demonstrated that the velum is actively lowered during normal speech, primarily by activity of the palatoglossus muscle. A large amount of palatoglossus activity is required for a nasal consonant. It is logical to assume that the nasalization of a preceding vowel could be a consequence of neuromuscular anticipation. The palatoglossus muscle would begin to contract before the nasal consonant, in anticipation of the activity for the consonant. However, the palatoglossus activity would be antagonistic to a gesture lowering the body of the tongue for low vowels, and it is the low vowels which are most often found to be nasalized before nasal consonants (Schane 1969). Neuromuscular anticipation cannot easily explain this apparently natural phenomenon either. Vowel nasalization may be a purely articulatory assimilation, however. Lubker and Moll (1965) and Fritzell (1969) have shown that the velum is opened further for nasalized low vowels than for high nasalized vowels. This is a result of the mechanical linkage between the tongue body and the soft palate. The production of a low vowel will tend to pull the soft palate downwards, making vowel nasalization more likely for a low vowel than for mid or high vowels. Lightner (1970) has presented a great deal of data which indicates that vowel nasalization in a number of languages is not a purely phonetic process. In Polish, for example, vowel nasalization is dependent upon the item which follows the nasal consonant. The vowel becomes nasalized before nasal consonants which are followed by a word boundary or by a continuant consonant (/s/).

There are a number of natural phonological rules which cannot be explained in terms of neuromuscular processes. Many languages have a rule such as (9) which makes nasals homorganic with the following obstruent.

(9) $$\begin{bmatrix} \text{C} \\ + \text{ nasal} \end{bmatrix} \rightarrow \begin{bmatrix} \alpha \text{ anterior} \\ \beta \text{ coronal} \end{bmatrix} \; / \; \underline{\qquad} \begin{bmatrix} - \text{ sonorant} \\ \alpha \text{ anterior} \\ \beta \text{ coronal} \end{bmatrix}$$

This is clearly a very natural rule, but a shift from [+ anterior, - coronal] (/m/) to [- anterior, - coronal] (/ŋ/) cannot be explained in terms of anticipation at the neuromuscular level. Quite obviously different spatial targets must be assigned to these nasal consonants in this case.

An extensive investigation and discussion of natural rules in phonology is far beyond the scope of this experimentally-oriented study. Nevertheless, the muscular phenomena discussed might eventually provide an explanation for some of the naturalness found in phonological rules. Other types of natural rules may eventually be accounted for by other phenomena. Schane (1969) discusses natural rules which serve to make segments perceptually distinct, and rule strategies which conspire to achieve a basic CV pattern. These phenomena cannot obviously be explained in terms of neuromuscular patterns of activity.

CONCLUSIONS

One of the basic results of this study has been the demonstration that the articulatory configurations of the tongue during the production of phonologically similar segments are not principally a result of "missed targets" or "articulatory undershoot." The different tongue positions for a fronted and non-fronted /k/ are produced by different neuromuscular patterns. These findings are in basic agreement with most other electromyographic studies of normal speech articulation, that is, many "allophonic" differences in articulatory gestures are reflected in the motor commands to the muscles. In previous studies, however, these differences in the motor command patterns have been ascribed to peripheral neurophysiological feedback processes such as the gamma efferent system. While we do not deny that such systems may play an important role in the control of speech, our investigations indicate that its role may be somewhat more limited than has been supposed in previous models.

Furthermore we have concluded, on the basis of a limited amount of data, that a *target command* model of speech production offers a good possibility of relating linguistic units such as phonological/phonetic features to physical phonetic events such as muscle motor commands. It is in the area of relations between linguistic units such as features and physical phonetic data that deserves the most intensive investigation in the future. The problems encountered in such investigations will hopefully lead to more sophisticated models of speech production, and a better understanding of the procedures by which detailed phonetic specification can be assigned to phonological features.

BIBLIOGRAPHY

Abd-el-Malek, S. (1939), "Observations on the morphology of the human tongue," *J. Anat. 73*, 201.

Amerman, J.D., Daniloff, R., and Moll, K.L. (1970), "Lip and jaw coarticulation for the phoneme /æ/," *JSHR 13*, 147-161.

Bach, E. (1968), "Two proposals concerning the simplicity metric in phonology," *Glossa 2*, 128-149.

Bach, E. and Harms, R. (1969), "How languages get crazy rules," paper presented at the 1969 UCLA Conference on Historical Linguistics.

Barker, D. (1966), "The motor innervation of the mammalian muscle spindle," in R. Granit (ed.), *Muscle Afferents and Motor Control*, 51-58.

Basmajian, J.V. and Stecko, G. (1962), "A new bipolar electrode for electromyography," *J. Appl. Physiol. 17*, 849.

Blom, S. (1960), "Afferent influences on tongue muscle activity," *Acta Physiol. Scand. 49*, 1-97.

Bole, C.T. II and Lessler, M.A. (1966), "Electromyography of the genioglossus muscles in man," *J. Appl. Physiol. 21*, 1695-1698.

Campbell, E. (1968), "The respiratory muscles," *Annals of the New York Academy of Sciences 155*, 135-140.

Caraceni, T. and Zibordi, F. (1966), "L'indagine elettromiografica in otorinolaringologia," *Arch. Ital. Otolog., Supplement 48*.

Chomsky, N. (1964), *Current Issues in Linguistic Theory*, The Hague: Mouton.

Chomsky, N. and Halle, M. (1968), *The Sound Pattern of English*, New York: Harper and Row.

Cooper, F. (1966), "Describing the speech process in motor command terms," Status Report on Speech Research , Haskins Laboratories.

Cooper, S. (1953), "Muscle spindles in the intrinsic muscles of the human tongue," *J. Physiol. 122*, 193-202.

----- (1960), "Muscle spindles and other muscle receptors," in G.H. Bourne (ed.), *The Structure and Function of Muscle, Vol. I*, 381-420.

Cosi, V. and Tonali, P. (1966), "Studio elettromiografico dei muscoli linguali, con particolare riguardo alla stimolazione del nervo ipoglosso, in soggetti normali," *Rivista di Neurobiologia 12*, 198-205.

Cunningham, D.P. and Basmajian, J.V. (1968), "Electromyography of tongue muscles," *Proc. Can. Fed. Biol. Soc. 11*, 15.

Dabelow, R. (1951), "Preliminary studies of the tongue as a functional system," *Morph. Jahrb. 91*, 33 [in German].

DeClerk, J.L., Landa, L.S., Phyfe, D.L., and Silverman, S.I. (1965), "Cinefluoragraphy of the vocal tract," *Proc. Vth Int. Congr. Acoustics*, Liege.

DeLattre, P. (1967), *The General Phonetic Characteristics of Languages*, mimeograph, U.C. Santa Barbara.

Drachman, G. (1969), "Adaptation in the speech tract," *Papers from the 5th Regional Meeting--Chicago Ling. Soc.*, 314-329.

Draper, M.H., Ladefoged, P., and Whitteridge, D. (1958), "Respiratory muscles in speech," *JSHR 2*, 16-27.

Eccles, J.C. (1966), "Functional organization of the cerebellum in relation to its role in motor control," in R. Granit (ed.), *Muscular Afferents and Motor Control*, New York: Wiley and Sons, 19-36.

----- (1967), "The way in which the cerebellum processes sensory information from muscle," in M.D. Yahr and D.P. Pupura (eds.), *Neurophysiological Basis of Normal and Abnormal Motor Activities*, 379-414.

----- (1969), "The dynamic loop hypothesis of movement control," in K.N. Leibovic (ed.), *Information Processing in the Nervous System*, New York: Springer-Verlag, 245-269.

Eccles, J.C., Ito, M., and Szentágothai, J. (1967), *The Cerebellum as a Neuronal Machine*, New York: Springer-Verlag.

Evarts, E.V. (1967), "Representation of movements and muscle by pyramidal tract neurons of the precentral motor cortex," in M.D. Yahr and D.P. Pupura (eds.), *Neurophysiological Basis of Normal and Abnormal Motor Activities*, New York, 215-253.

Eyzaguirre, C.E. (1969), *Physiology of the Nervous System*, Chicago: Year Book Medical Publishers.

Faaborg-Andersen, K. and Vennard, W. (1964), "Electromyography of extrinsic laryngeal muscles during phonation of different vowels," *Ann. Otol. 73*, 248-254.

Fant, G. (1960), *Acoustic Theory of Speech Production*, The Hague: Mouton.

----- (1969), "Distinctive features and phonetic dimensions," *Speech Trans. Lab. Quart. Prog. and Status Report 2-3/1969*, 1-18.

Flanagan, J.L. (1965), *Speech Analysis, Synthesis, and Perception*, New York: Academic Press.

Flanagan, J.L., Coker, C.H., Rabiner, L.R., Schafer, R.W., and Umeda, N. (1970), *IEEE Spectrum*, October 1970, 22-45.

Fritzell, B. (1969), *The Velopharyngeal muscles in speech*, Suppl. 250 to *Acta Otol.*

Fromkin, V.A. (1966), "Neuromuscular specification of linguistic units," *Language and Speech 9*, 170-199.

----- (1968), "Speculations on performance models," *J. of Ling. 4*, 47-68.

----- (1970), "Tips of the slung -- or -- to err is human," *UCLA Working Papers in Phonetics 14*, 40-79.

Fromkin, V.A. and Ladefoged, P. (1966), "Electromyography in speech research," *Phonetica 15*, 219-242.

Goldstein, M.A. (1940), "Speech without a tongue," *JSHD 5*, 65-69.

Granit, R. (1955), *Receptors and Sensory Perception*, New Haven: Yale Univ. Press.

----- (ed.) (1966), *Muscular Afferents and Motor Control*, New York: Wiley and Sons.

Granit, R., Holmgren, B., and Merton, P.A. (1955), "Two routes for the excitation of muscle and their subservience to the cerebellum," *J. Physiol. 130*, 213-224.

Harris, K.S., Lysaught, G.F., and Schvey, M.M. (1965), "Some aspects of the production of oral and nasal labial stops," *Lang. and Speech 8*, 135-147.

Harris, K.S., Gary, T., Sholes, G.N., and Lieberman, P. (1969), "Some stress effects on electromyographic measures of consonant articulation," *Status Rep. on Speech Research 13/14,* Haskins Laboratories, New York City.

Harshman, R. and Ladefoged, P. (1967), "The LINC-8 computer in speech research," *UCLA Working Papers in Phonetics 7,* 57-68.

Henke, W. (1966), *Dynamic Articulatory Model of Speech Production Using Computer Simulation.* Unpublished dissertation, Massachusetts Institute of Technology, September 1966. Summarized in Speech Communication Research Final Report, Massachusetts Institute of Technology Research Laboratory of Electronics, July, 1966.

Henneman, E. and Olson, C.B. (1965), "Relations between structure and function in the design of skeletal muscle," *J. Neurophysiol. 28,* 581-598.

Henneman, E., Somjen, G., and Carpenter, D.O. (1965), "Excitability and inhibitability of motoneurons of different sizes," *J. Neurophysiol. 28,* 599-620.

Hiki, S., Ratcliffe, R., Hubler, S., and Metevelis, P. (1968), "Notes on LASS circuitry," *UCLA Working Papers in Phonetics 10,* 12-41.

Hirano, M., Koike, Y., and von Leden, H. (1967), "The sternohyoid muscle during phonation," *Acta Otolaryngologica 64,* 500-507.

Hirano, M. and Ohala, J. (1969), "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles," *JSHR 12,* 362-373.

Houde, R.A. (1967), *A Study of Tongue Body Motion During Selected Speech Sounds.* Unpublished Ph.D. dissertation, Univ. of Michigan.

Hubler, S. (1967), "A high input impedance electromyography preamplifier," *UCLA Working Papers in Phonetics 7,* 25-34.

Hyman, L. (1970), "How concrete is phonology?," *Lang. 46,* 58-76.

Jung, R. and Hassler, R. (1960), "The extrapyramidal motor system," in Field et al. (eds.), *Handbook of Physiology,* Chapter 35, 863-928, Washington: American Physiological Society.

Kelsey, C.A., Ewanowski, S.J., Hixon, T.J., and Minifie, F.D. (1968), "Determination of lateral pharyngeal wall motion during connected speech by use of pulsed ultrasound," *Science 161,* 1259-1260.

Kent, R.D. and Moll, K.L. (1969), "Vocal-tract characteristics of the stop cognates," *JASA 46,* 1549-1555.

Kim, C.-W. (1966), *The Linguistic Specification of Speech, UCLA Working Papers in Phonetics 5.*

Kozhevnikov, V.A. and Chistovich, L.A. (1965), *Speech: Articulation and Perception,* U.S. Dept. of Commerce translation, JPRS 30:543.

Ladefoged, P. (1964), "Some possibilities in speech synthesis," *Language and Speech 7,* 205-214.

----- (1964b), *A Phonetic Study of West African Languages.* London: Cambridge University Press.

----- (1967a), *Three Areas of Experimental Phonetics,* London: Oxford University Press.

----- (1967b), *Linguistic Phonetics, UCLA Working Papers in Phonetics 6.*

----- (1970), "The phonetic framework of generative phonology," *Working Papers in Phonetics 14,* 25-33.

Ladefoged, P. and Fromkin, V. (1968), "Experiments on competence and performance," *IEEE Trans. Audio Electroacoustics AU-16,* 130-136.

Laver, J. (1968), "Phonetics and the brain," *Work in Progress No. 2,* Edinburgh University, 63-75.

Law, M.E. (1954), "Lingual proprioception in pig, dog and cat," *Nature 174,* 1107-1108.

Lee, C. (personal communication)

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967), "Perception of the speech code," *Psychol. Rev. 74,* 431-461.

Lieberman, P. (1965), "On the acoustic basis of the perception of intonation by linguists," *Word 21,* 40-54.

Lightner, T. (1970), "Why and how does vowel nasalization take place?," *Papers in Ling. 2,* 179-226.

Lubker, J.F. and Moll, K.L. (1965), "Simultaneous oral-nasal air flow measurements and cinefluorographic observations during speech production," *The Cleft Palate Journal, Vol. 2, No. 3*, 257-272.

Lucas Keene, M.F. (1961), "Muscle spindles in the human laryngeal muscles," *J. Anat. 95*, 25-29.

McCawley, J.D. (1968), *The Phonological Component of a Grammar of Japanese*, The Hague: Mouton.

MacConaill, M.A. and Basmajian, J.V. (1969), *Muscles and Movements*, Baltimore: Williams and Wilkins.

MacNeilage, P.F. (1970), "Motor control of serial ordering of speech," *Psychol. Rev. 77*, 182-196.

MacNeilage, P.F. and DeClerk, J.L. (1969), "On the motor control of coarticulation in CVC monosyllables," *JASA 45*, 1217-1233.

MacNeilage, P.F. and Sholes, G.N. (1964), "An electromyographic study of the tongue during vowel production," *JSHR 7*, 209-232.

Mashiko, N. (1960), "Studies on clinical electromyography (II), electromyograms on tongue muscles," *Folia Psychiatrica et Neurologica Japonica 14*, 28-41

Matthews, P.B.C. (1964), "Muscle spindles and their motor control," *Physiol. Reviews 44*, 219-288.

Mattingly, I. (1967), "Speech synthesis by rule as a research technique," *Status Report on Speech Research SR-10*, Haskins Laboratories, New York City.

Moll, K.L. (1960), "Cinefluorographic techniques in speech research," *JSHR 3*, 227-241.

Ohala, J. (1970), *Aspects of the Control and Production of Speech, UCLA Working Papers in Phonetics 15.*

Ohala, J. and Hirano, M. (1967), "Control mechanism for the sequencing of neuromuscular events in speech," paper presented at the 1967 Conference on Speech Communication and Processing, M.I.T.

Öhman, S. (1966), "New methods for averaging EMG records," *Speech Trans. Lab. Quart. Prog. and Status Report 1/1966*, 5-8.

----- (1967a), "Numerical model of coarticulation," *JASA 41*, 310-320.

----- (1967b), "Peripheral motor commands in labial articulation," *Speech Trans. Lab. Quart. Prog. and Status Report 4/1967*, 30-63.

Öhman, S., Persson, A., and Leanderson, R. (1967), "Speech production at the neuromuscular level," *Speech Trans. Lab. Quart. Prog. and Status Report 2-3/1967*, 15-19.

Perkell, J.S. (1969), *The Physiology of Speech Production*, Cambridge: M.I.T. Press.

Pernkopf, E. (1963), *Atlas of Topographical and Applied Human Anatomy, Vol. I: The Head and Neck*, H. Ferner (ed.), translated by H. Monsen, Philadelphia: W.B. Saunders.

Porter, R. (1965), "Synaptic potentials in hypoglossal motoneurons," *J. Physiol. 180*, 209-224.

----- (1966), "Lingual mechanoreceptors activated by muscle twitch," *J. Physiol. 183*, 101-111.

Postal, P. (1968), *Aspects of Phonological Theory*, New York: Harper and Row.

Rice, L. (forthcoming in *UCLA Working Papers in Phonetics 18*)

Ringel, R.L. and Steer, M.D. (1963), "Some effects of tactile and auditory alterations on speech output," *JSHR 6*, 369-378.

Rosen, G. (1960), "Dynamic analog speech synthesizer," Technical Report 353, Res. Lab. Electr., M.I.T.

Schachter, P. (1969), "Natural assimilation rules in Akan," *IJAL 35*, 342-356.

Schachter, P. and Fromkin, V. (1968), *A Phonology of Akan: Akuapem, Asante, and Fante*, *UCLA Working Papers in Phonetics 9*.

Schane, S. (1969), "Natural rules in phonology," paper read at the 1969 UCLA Conference on Historical Linguistics.

Sears, T.A. and Newsom Davis, J. (1968), "The control of respiratory muscles during voluntary breathing," *Annals of the New York Academy of Sciences 155*, 183-190.

Spalteholz, W. (1933), *Hand-atlas of Human Anatomy* (7th English Edition), translated by L.F. Barker, Philadelphia: Lippincott.

Stanley, R. (1967), "Redundancy rules in phonology," *Lang. 43,* 393-436.

Stevens, K.N. and Halle, M. (1967), "Remarks on analysis-by-synthesis and distinctive features," in W. Wathen-Dunn (ed.), *Models for the Perception of Speech and Visual Form,* Cambridge: M.I.T. Press.

Stevens, K. and House, A. (1955), "Development of a quantitative description of vowel articulation," *JASA 27,* 484-493.

----- (1961), "An acoustical theory of vowel production and some of its implications," *JSHR 4,* 303-320.

Stevens, K.N. Kasowski, S., and Fant, G. (1953), "An electrical analog of the vocal tract," *JASA 25,* 734-865.

Strong, L.H. (1956), "Muscle fibers of the tongue functional in constant production," *Anat. Record 126,* 61.

Szentagothai, J. (1948), "Anatomical considerations of monosynaptic reflex arcs," *J. Neurophysiol. 11,* 445-453.

Tatham, M.A.A. (1969a), "Control of the muscles in speech," University of Essex Language Centre, *Occasional Papers No. 3,* 23-40.

----- (1969b), "Classifying allophones, University of Essex Language Centre, *Occasional Papers No. 3,* 14-22.

----- (1970), "Articulatory speech synthesis by rule: implementation of a theory of speech production," ditto, University of Essex.

Tatham, M.A.A. and Morton, K. (1969), "Some electromyographic data towards a model of speech production." *Language and Speech.*

Tokizane, T. and Shimazu, H. (1964), *Functional Differentiation of Human Skeletal Muscle,* Tokyo: Univ. of Tokyo Press.

Umeda, N. and Teranishi, R. (1966), "Speech synthesis by acoustic vocal tract model," *J. Acoust, Soc. Japan 12,* 195-203.

Van Riper, C. and Irwin, J.V. (1958), *Voice and Articulation,* Englewood Cliffs, New Jersey: Prentice-Hall.

Voss, H. (1956), "Zahl und Anordnung der Muskelspindeln in den oberen Zungenbein muskeln, im M. trapezius und M. latissimus dorsi," *Anat. Anz. 103,* 443-446.

Wickelgren, W.A. (1965), "Acoustic similarity and intrusion errors in short-term memory," *J. Exp. Psychol. 70,* 102-108.

Wickelgren, W.A. (1969), "Context-sensitive coding, associative memory, and serial order in (speech) behavior," *Psychol. Rev. 76,* 1-15.