DOCUMENT RESUME

ED 064 971 FL 003 201

AUTHOR	Ohala, John J.
TITLE	Aspects of the Control and Production of Speech.
INSTITUTION	California Univ., Los Angeles. Dept. of Linguistics.
SPONS AGENCY	National Institutes of Health (DHEW), Bethesda, Md.; Office of Naval Research, Washington, D.C.
PUB DATE	Apr 70
NOTE	195p.; In "Working Papers in Phonetics 15," Revised version of doctoral dissertation
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Acoustic Phonetics; Artificial Speech; Auditory Perception; Cognitive Processes; Distinctive Features; Experiments; Intonation; Language Patterns; *Language Research; Linguistic Theory; Neurolinguistics; *Phonetics; Phonology; *Physiology; Psycholinguistics; Psychology; *Research Methodology; *Speech; Speech Pathology

ABSTRACT

The dominant trend in phonetics today--due to a large extent to generative phonology--is to discover the brain mechanisms underlying the observed behavior in speech. Among other things there is interest in attempting to find out how motor programs are stored latently, selected, activated into muscular contractions, controlled, and tailored for optimum communication. Chapter 1 discusses research thought and methodology in phonetics and expresses the need for constant empirical feedback in all stages of the development of models of these processes. Chapter 2 attempts to shed light on the mechanisms speakers use to control the fundamental frequency of phonation in speech. Chapter 3 considers arguments, evidence, and experimental techniques relevant to discovering certain possible brain mechanisms underlying observed speech behavior. Two issues are covered: the possible role of feedback in speech and how the timing of gestures in speech is controlled. (Author/VM)

Aspects of the control and production of speech

John J. Ohala

Working papers in phonetics 15

April 1970

University of California, Los Angeles

Acknowledgements

The present work is a revised version of a doctoral dissertation of the same title submitted in May 1969 to the Department of Linguistics at the University of California at Los Angeles. I would appreciate comments and criticism on this work from interested readers.

In writing this work and in conducting the research reported in it, I have benefited from the encouragement and assistance given me by those who have served on my doctoral committee, Victoria Fromkin, Robert Wilson, Terence Moore, Donald MacKay, Douglas Junge, Robert Stockwell, and, especially, my chairman and mentor, Peter Ladefoged. In addition I have profited from exchanges of ideas with my colleagues in the Phonetics Laboratory at UCLA and in the Research Institute of Logopedics and Phoniatrics at the University of Tokyo, in particular, George Allen, Osamu Fujimura, Robert Greenberg, Richard Harshman, Shizuo Hiki, Minoru Hirano, Hajime Hirose, Stanley Hubler, Chin-Woo Kim, Mona Lindau, and Timothy Smith. Kerstin Hadding and Ralph Vanderslice have also provided me with valuable detailed comments on earlier versions. The technicians in the Phonetics Laboratory, especially Willie Martin, have generously given of their time in assisting me with the various experiments reported here, and Jeanne Yamane, Renee Wellin, and Julie Haaker have skillfully converted my largely illegible manuscripts into clean typescript. Leon Jacobson helped immensely with the bibliography as well as with other mechanics of the manuscript. To my wife, Manjari, I owe special thanks for her help in all stages of the preparation of this work, from assistance in experiments, to proofreading, to providing helpful criticism. To all of these, my friends and colleagues, I can only express my regret that this work represents such a poor return on their rich investment of interest, encouragement, and help.

John Ohala
Department of Linguistics
University of California, Berkeley
April, 1970

(Additional revisions are incorporated in this Xerox copy.

J.J.O. December 1970)

This research was supported in part by a United States Public Health Services (NIH) grant, NB-04595, and in part by an Office of Naval Research contract, NR-049-226.

Abstract

Chapter One: The dominant trend in phonetics today -- due to a large extent to Generative Phonology -- is to discover the brain mechanisms underlying the observed behavior in speech. Among other things there is interest in attempting to find out how motor programs are stored latently, selected, activated into muscular contractions, controlled, and tailored for optimum communication. Constant empirical feedback is needed in all stages of the development of models of these processes.

Chapter Two: It has traditionally been assumed that the modulation of the fundamental frequency of phonation in speech is predominantly controlled by the laryngeal muscles. However, Lieberman (*Intonation, Perception, and Language*, M.I.T. Press, 1967) challenges this view and proposes that the fall in pitch at the end of sentences and the rise in pitch on emphasized words are both determined mainly by variations in the subglottal pressure, and that except for the rise in pitch at the end of yes-no questions, the laryngeal muscles are usually not involved.

To test this hypothesis the activity of some of the laryngeal muscles was sampled electromyographically in five subjects as they spoke sentences with a variety of intonation contours. It was found for all subjects that the laryngeal muscles participate actively in modulating pitch, in particular the cricothyroid and lateral cricoarytenoid muscles are active in raising pitch and the sternohyoid muscle in lowering pitch. This is true no matter where the pitch modulation occurs in a sentence and no matter what grammatical or semantic entity is manifested by the pitch change. The effect of subglottal pressure changes on pitch was calibrated by the familiar "push-on-the-chest" technique and was found to be about 2-3 Hz/cm. water for one subject in the pitch range used in speech. This was too small a value to account for most of any given pitch change recorded in speech.

The evidence did not support Lieberman's theory of intonation but re-affirmed the traditional view that the speaker can and does "program" his larynx to execute any intonational pattern he desires. Additional evidence adduced by Lieberman to show that certain aspects of intonation were innately determined and hence universal were critically examined and found to be weak.

In addition, recent claims made about the perception of "stress" by trained linguists were examined in the light of recent experimental findings. Some explanations are proposed for the perceptual origin of the multiple levels of stress in English.

Chapter Three: Arguments and evidence from jaw movements were presented demonstrating the possibility and likelihood of the use of rapid kinesthetic feedback in speech. Also two experiments are reported relevant to the question of how the gestures in speech are sequenced. Evidence could not be found for the claim that an independent time schedule -- isochronic or not -- underlies speech utterances, but much more research is needed on this point.

... if it is to be interesting, the description of each language must also be testable; and the possibility of making a sufficient test must be inherent in the underlying theory.

— Peter Ladefoged (1967b)

Chapter 1

At a special session of the 1967 Conference on Speech Communication and Processing, held at M.I.T., Kenneth Stevens expressed the opinion that speech research was today at about the same stage that chemistry was *before* the periodic table of elements had been discovered. Unfortunately for the field of speech, the analogy is quite appropriate. As was the case with chemistry in the first half of the 19th century, speech research does not seem to be able to make all of its disparate pieces of data fit together into a general pattern of any sort. Typically it is unable to generalize beyond the available data and cannot tell what kind of experiments are worth doing.

However, this is not to deny that progress has been made in the field over the past 100-150 years, but mainly in the accumulation of a wider variety of high-quality data. Much more is known about the behavior of the vocal tract and the acoustic properties of speech such that it is now possible to synthesize speech with some success. And there has been some limited advancement on the problem of automatic speech recognition.

Perhaps the most telling sign of progress in this area is the shift in attention from the purely peripheral aspects of speech to the more central, neurological substrate of speech. The periphery, i.e., where sounds are produced, acoustically and neuromuscularly, is still being studied, of course, but more and more for what it will reveal about the brain mechanisms underlying speech. Quite sensibly it is believed that if we could find out what is happening inside the "box" (the speaker's head) we might be able to make more sense of the sounds that are emitted from it and of the perceptual judgments made by it. This has been the explicit or implicit strategy of many researchers in phonetics/phonology, one of the disciplines in speech research. In this chapter I propose to discuss how this strategy should and should not be pursued within phonetics/phonology.

Of the many disciplines included in speech research, e.g., communications engineering, oto-laryngology, speech pathology, psychology, physiology, and phonetics/phonology (although the dividing lines between these disciplines are fast disappearing), phonetics/phonology has played and still does play a rather central role

in the field of speech research, and thus possibly links the success or failure of the whole field with its own achievements. Thus, wisely or not, most of the other disciplines have freely adopted the terms and concepts of phonetics in their own specialized work. And so, terms such as "phoneme", "syllable", "vowel and consonant", "phone" and "allophone" seem to have a more secure, unquestioned position in the writings of communications engineers than they do in those of phoneticians.

Phonetics/phonology has contributed to and shared in whatever success has been enjoyed by all of speech research, and in the process has experienced the same shifting of its goals. Although attention has never shifted from its original tasks of providing accurate physiological and acoustic descriptions of sounds and sound patterns in the world's languages or of providing explanations for sound changes and sound alternations in languages, there is now increasing concern for discovering the brain mechanisms underlying the observed speech behavior. Compare, for example, the topics discussed at the first and second International Congresses of Phonetic Sciences with those at the fourth and fifth: now we may expect a large number of papers dealing with language and speech in general rather than detailed studies of particular languages.

However phonetics has always had a handicap that the other disciplines did not have, namely it was not wholly convinced that it could or should be an experimental science of the same rigor as physiology or chemistry. This is undoubtedly due to its origin as a branch of philology, a discipline in which it is impossible to employ laboratory experiments. Perhaps phonetics started out as a serious science, but meeting many initial failures, retreated back into the safe, less frustrating world of *a priori-ism* typical of philosophy and literary criticism. It is possible to find among early phonetic studies some well-intentioned if primitive attempts to subject hypotheses about speech to experimental verification. Grandgent (1890) in the same article in which he reports on a creditable attempt to discover the true tongue positions for certain vowels (by sticking wire measuring "probes" into his mouth), finds it necessary to warn against unwholesome, unscientific attitudes in the field.* However we know now that

---

* Grandgent's complaint was a universal and timeless one: he asserted that one of the reasons why phonetics had not made much progress even in those days was

> ... the well-nigh irresistable tendency to construct theories on insufficient data. Many investigators have, I fear, designed their system first, and then pared off the toes and heels of their facts to make them fit the symmetrical slipper into which they were to be thrust.

even Grandgent's findings were not very accurate in spite of his care and objective attitude, and perhaps it was the resulting endless disputes and the frustration experienced from repeated failure at such experiments that caused some phoneticians to be disapproving of instruments. Thus, Sweet (1911) in the entry "Phonetics" for the 11th edition of the *Encyclopedia Britannica* briefly outlines the accomplishments of instrumental phonetics but comments

> There has been great discrepancy between the results obtained by different observers, and many results which were at first received with implicit confidence for their supposed rigorously scientific and objective character have been found to be worthless.

Sweet wrote these words in his last years and at the end of his long career in phonetics (he was 65 years old when they were published and he died the following year in 1912). We can therefore guess that there may be more disappointment than bitterness underlying his final haughty write-off of instrumental phonetics in the following passage.

> The claims of instrumental phonetics have been so prominently brought forward of late years that they can no longer be ignored even by the most conservative of the older generation of phoneticians. But it is possible to go too far the other way. Some of the younger generation seem to think that the instrumental methods have superseded the natural ones in the same way as the Arabic superseded the Roman numerals. This assumption has had disastrous results. It cannot be too often repeated that instrumental phonetics is, strictly speaking, not phonetics at all. It is only a help: it only supplies materials which are useless till they have been tested and accepted from the linguistic phonetician's point of view. The final arbiter in all phonetic questions is the trained ear of a practical phonetician: differences which cannot be perceived must -- or at least may be -- ignored; what contradicts the trained ear cannot be accepted.

Considering the trials that instrumental phonetics suffered in its early years it is more appropriate that we read these comments by Sweet with compassion rather than disapproval. However the situation is different now and it is very surprising to find Sweet's statements endorsed by modern phoneticians -- even by those who are also instrumentalists. Ladefoged (1964) quoted the above passage by Sweet and added:

> For those of us who are not as skilled as Sweet, instrumental phonetics may be a very powerful aid and of great use in providing objective records on the basis of which we may verify or amend our subjective impressions. But even the most extensive array of instruments can never be a substitute for the linguist's accurate observation and imitation of an informant.

This passage in turn was quoted approvingly by Stevens and Halle (1967) in support of their claims on the abstract and mentalistic character of phonetic units. Ladefoged, however, points out that the statement appeared in a book dealing with the phonetic description of the phonological contrasts in a number of languages studied in the field. It was not meant to imply that a linguist's observations were preferable to experimental studies in other circumstances (Ladefoged 1969, personal communication).

However, more debilitating to the field than scorn of instrumental studies was the disregard of normal scientific procedure, particularly with respect to testing one's hypotheses. Testing need not involve fancy instruments. It amounts merely to refining one's observations by constructing or finding a situation in which the phenomenon or relation one predicts will be free to appear or not appear -- free, that is, from irrelevant influences. Introspection and intuition are, of course, perfectly legitimate sources for initial observations. No one really cares how a scientist gets his ideas for hypotheses; but people do care whether or not they are subsequently tested, i.e., submitted to potential falsification. One cannot terminate one's study with introspection because it is difficult, if not impossible altogether, to replicate it, and because the results are too susceptible of being rigged -- even unconsciously. As Grandgent (1890) noted, in arguing against introspective experiments in phonetics,

> ... sensation is ... uncertain; for feeling depends far less on the actual movements of the organs than on the preconceived idea in the observer's mind.

As a result of much flagrant disregard for testing even of its most fundamental ideas, phonetics now has an abundance of terms which have only an imperfectly understood meaning, e.g., syllable, stress, tense versus lax, etc.

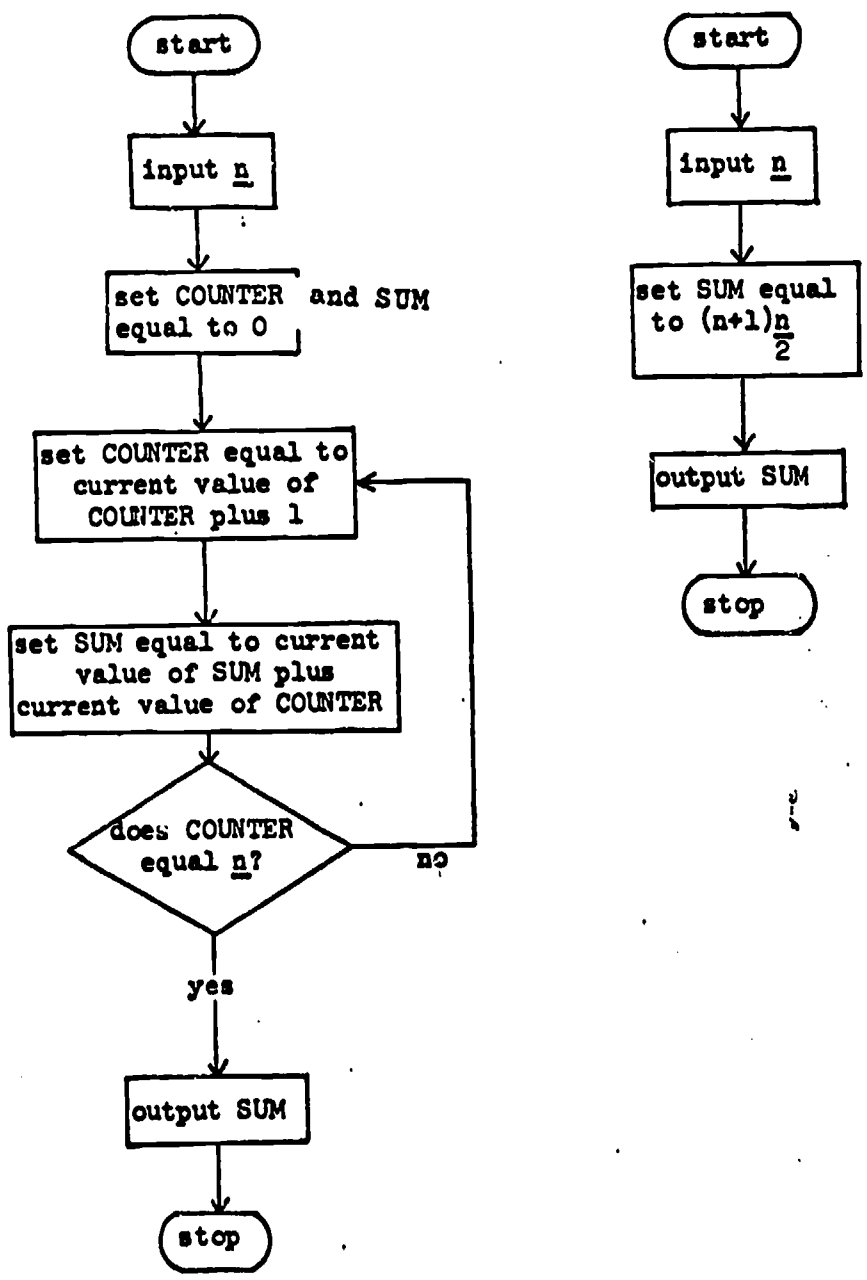
This attitude is still with us and generative phonology, as put forth in the writings of Chomsky, Halle, Postal, Lightner, etc., for all the good it has done the field by re-channelling interest along more profitable lines, remains one of the most ardent proponents of this scientifically anachronistic mode of thought. Of course, this school does frequently allow that certain issues are "empirical matters," i.e., need to be tested, (although what assumption of fact and statement of hypothesis is not an empirical matter?), but one may question how long a discipline interested in "the nature of mental processes" (Chomsky and Halle 1968, p. viii) may proceed without actually performing any convincing experiments involving a human and his mental processes. The only means of verification the generative phonologists seem to admit is testing whether or not their system, primitive units plus rules, is capable of "capturing" certain linguistic generalizations. But before this can lead to information on

mental processes, two more steps must be taken. As Zimmer (1968) has recently pointed out,

> It seems to be occasionally assumed that the very fact that a [linguistic] regularity can be stated suffices to justify the inference that it has some kind of psychological reality, but there is surely nothing necessary about this assumption. One might equally well assume that someone who learns the sequence of numbers 1, 5, 19, 65, 211, 665 must necessarily know the formula which relates them (namely that the nth member of the series equals 3^n-2^n). Of course if this person could not only repeat the sequence correctly, but also continued it on his own with 2059, that would be evidence that he knows the formula in question, but it is just evidence of such a conclusive nature that is lacking in the case of some of the regularities we find in languages.

After writing a rule, the first step, then, is to see if it represents a linguistic process that is productive in the native speaker. Testing this point on some phonological rules in Turkish, Zimmer found that "speakers may well be only very imperfectly aware of regularities that do not fall within the system of fully productive rules of their language." Ladefoged and Fromkin (1968), however, in an experiment in which subjects were presented with written nonsense words such as *sublane*, *sublantory*, *sublanation*, etc. and were asked to transcribe the words phonetically, found that subjects could be phonologically productive along the lines set down in certain rules of English phonology as formulated by Chomsky and Halle.

But the next step after this is as important as it is difficult. This is to ascertain how close the rules, as written, match the process used by the native speaker. Consider another series of numbers, 1, 3, 6, 10, 15, 21, 28, and assume that the subject can produce these and all of the succeeding terms, 36, 45, etc., i.e., the nth term of the series equals $\sum_{i=1}^{n} (i)$. At least two different algorithms, shown in Figure 1, can be used to arrive at the correct value for the nth term. Knowing the second algorithm a bright high-school student, given an n of 10,000,000,000 could, within a fraction of one minute, come up with the correct value for the nth term of the series, 50,000,000,005,000,000,000, a feat which would better many computers using the first algorithm. If we "psycho-mathematicians" only knew of the first algorithm we would draw some fantastic and erroneous conclusions about the computing ability of the human brain if we assumed the human were also using the first algorithm. Thus, even if we grant for the moment that the transformational cycle in phonology describes a process that is productive in speakers of English (although see



Algorithm 1 Algorithm 2

Figure 1. Two algorithms which take the same input and give the same output.

counter-arguments in Chapter 2, below), need it follow that the speaker is using an algorithm as complex as the transformational cycle? It does *not* follow, for the reasons given above and yet it is such a claim which Chomsky and Halle use to attribute innate linguistic structure to human brains, there being, supposedly, insufficient data or time to learn such a complicated algorithm inductively. Clearly we are in need of many more tests.

But generative phonology has adopted an incredible myth about the function of tests, namely that they must always support the hypothesis tested, otherwise they are defective:

> ... operational* tests, just as explanatory theories, must meet the condition of correspondence to introspective judgement, if they are to be at all to the point.
> [Chomsky 1964, p. 80]

> ... If a test designed to demonstrate behavioral correlates for ... [linguistic notions such as nouns, phonemes, etc.] ... fails to yield the predicted results, one feels obligated to modify the test, not the theory.
> [Saporta 1965]

Would that the world were so lucky as to have a hard core of clearly and unquestionably certain facts. If so, we could happily adopt the Chomskyan method in science. Unfortunately, history has seen more than one "intuitively obvious fact" overthrown by subsequent scientific investigation, e.g., the flatness of the earth, the motion of the sun through the heavens, that heavy objects fall faster than light ones, etc. Lacking omniscience, or divine revelation, scientists have decided that it is better not to throw out data if it conflicts with the posited hypothesis -- ideally, at least. They do however frequently present arguments against results that conflict with their hypotheses by claiming or demonstrating that the experiment is unreliable or invalid, or both. Unless scientists have good reason to discredit an experiment (e.g., previous experiments having already countered the revolutionary claim), they usually do not simply ignore its results.**

---

* It is not clear whether or not "operational" here means anything more than "involving operations"; if not, it is redundant in that all tests involve some operations.

** Cf., e.g., Helmholtz's account (1881) of the treatment accorded Goethe's "refutation" of Newton's theory that white light consisted of a mixture of all colors. Goethe's friends at first patiently pointed out to him how Newton's theory and experiments accounted for all of his objections, but when he persisted, they eventually just ignored him.

This procedure is followed even in cases where many previous experiments seem to have firmly established a certain theory. Witness, for example, the flood of articles offered as counterevidence to Husson's (1950) neurochronaxic theory of vocal cord vibration. Although it had been confidently held for over 200 years that the vibrations of the vocal cords were aerodynamic phenomena and not due to individual muscular contractions, a large amount of energy was taken to refute point-by-point most of Husson's innovative claims. The resulting research (by both sides) has benefitted the field immensely by enlarging the store of information about laryngeal physiology and in some cases by increasing the repertory of techniques available for studying the larynx. However, many of the theoretical points Chomsky and others espousing generative phonology hold to be so unquestionably true that they take supremacy over experimental results, e.g., the mental reality of phonemes, judgments of grammaticality, etc., do not even fall into the class of issues strongly supported by many previous experiments. All the more reason, I would think, for encouraging fellow linguists to start looking for ways of testing them. Instead Chomsky provides a discipline already having weak empirical foundations with neat rationalizations for why experiments should not be attempted.*

> ... at the present stage of the study of language, it seems rather obvious that the attempt to gain some insight into the range of data that we now have is likely to be far more fruitful than the attempt to make this data more firm, e.g., by tests for synonymy, grammaticalness, and the like. Operational criteria for these notions, were they available and correct, might soothe the scientific conscience, but how, in fact, would they advance our understanding of the nature of language or of the use and acquisition of language? [1964, p. 81]

How indeed? We can only find out if we try. If we do not try to firm up our hypotheses, if we do not try to get some empirical feedback on our initial guesses about the nature of the linguistic rules in the speaker's head, we are likely to develop a theory inflated with unreal and fantastic theoretical machinery exactly as happened in Cartesian cosmology. I believe Chomsky is completely wrong in the lesson he derives from the natural sciences when he considers the issue of

---

* One wonders whether or not this may also be part of the motivation for the development of that favorite subject of linguistic exegesis: the distinction between *competence* and *performance*. Thus it was asserted that the linguist's task was to discover and describe the native speaker's competence, i.e., what he knew about his language; however, tests or any sort of experimental evidence as to the presence or use of these rules were difficult or impossible to obtain due to "performance" factors which obscured the underlying competence. Linguists were assured however that statements about the speaker's competence could still safely be made in spite of these unsolved interferences by performance factors.

> ... whether the important feature of the successful sciences had been their search for insight or their concern for objectivity.
>
> ... a good case can be made for the view that the natural sciences have, by and large, sought objectivity primarily insofar as it is a tool for gaining insight (for providing phenomena that can suggest or test deep explanatory hypotheses).
>
> ... In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. [1965, p. 20]

Certainly chemists do not go around seeking the structure of any randomly selected substance just for the sake of being exact; of course they and all scientists seek insight. But the question is whether or not insight can be achieved *without* objectivity. If any lesson can be learned from the history of the "successful" sciences, the answer is a resounding NO. Chemistry only really started making rapid progress after Lavoisier, Priestly, Davy, Gay-Lussac, Berzelius, and others concerned with objectivity as well as "insight", had blessed the field. Similarly so with Physics and Newton, and with Physiology and Bernard.

It seems to be assumed that linguistics is the exception among sciences. It is assumed that since many of the posited phonological entities, such as the phoneme, any one of the Distinctive Features, or the transformational cycle, are all so abstract and mentalistic that evidence as to their psychological reality is necessarily hard to obtain given the current state of the art in psychological testing. Thus linguistics is apparently considered to be exempt from testing until the time comes when adequate tests are developed. This is false; the more abstract a posited entity is, the more likely it is to be wrong and thus the more it requires a test. Besides, as Northrop (1947) and others have pointed out, all the mature scientific disciplines posit unseeable entities, but this does not mean that there can be no evidence of their existence and operations. Halle (1964, p. 325) is completely unjustified in appealing to the acceptance of the electron, an unseeable entity, as the model for linguists' acceptance of the phoneme. The two cases are not at all comparable. When have linguists been treated to experiments as beautiful, as ingeniously devised, and as convincing as those performed by Hertz, Thompson, Wilson, Millikan, and others who have not only demonstrated the existence of the electron but many of its properties as well? Thus I would like to know exactly which "facts" Chomsky was referring to when he asserted

> Like most facts of interest and importance, [information about the speaker-hearer's competence] is neither presented for direct observation nor extractable from data by inductive procedures of any known sort. [1965, p. 18]

That the entities we are interested in are usually invisible is true, as mentioned above, but that they *cannot* be known by inductive procedures of any known sort has been refuted time and time again in science. Only if "facts of interest and importance" refer to theological propositions, perhaps, could this be justified. But I have been under the impression that linguistics is attempting to find out something about the nature of real mental processes. Is it possible to *know* something without being able to test it? Or, equivalently, is it possible to observe anything without it being possible to refine one's observations?

It is understandable that to the highly creative and imaginative minds that have given the world the intellectually satisfying system of Generative Phonology, the drudgery of experimentation, the painstaking putting of one empirical brick on top another is quite boring as indicated by the above quotes. But there really is no other way and there are many researchers in science who are convinced that their ingenuity and intellectual powers can be satisfyingly challenged in designing and running experiments as well as inventing theoretical systems which explain their experimental results. Otherwise we might as well revert back to Cartesian cosmology or Thomistic philosophy or the Sunday New York Times crossword puzzle, all three of which never fail to excite our esthetic and intellectual admiration but which advance the knowledge of the world not at all.

Chomsky on occasion seems willing to admit this:

> ... I do not think anybody actually working on language can doubt ... that sooner or later ... it is going to be necessary to discover conditions on theory constructions, coming presumably from experimental psychology or from neurology, which will resolve the alternatives that can be arrived at by the kind of speculative theory construction linguists can do on the basis of the data available to them. That is, there will come a point, no doubt, and I think in some areas of linguistics it may already have been reached, where one can set up alternative systems to explain quite a wide range of phenomena. One can think that this or that system is more elegant and much more deep than some other, but is it right?
>
> ... it seems to me that in phonology that point may have been reached. We can set up quite elegant theories of phonological structure that can explain quite a remarkable range of phenomena ... I think rather striking explanations can be proposed on the basis of theories which, although I think they are intellectually quite satisfying, have no evidence for them other than the fact that they explain quite a lot of phonetic data. Here, certainly, one hopes it will be possible to go beyond that, and you cannot go on beyond that on the basis of linguistics alone. [1967a, p. 100]

One could quibble over the implication that linguistics needs to be a non-experimental discipline and that it must look to other fields to do its testing or the suggestion that experimentation is required only at some critical stage in the progress of a theory rather than being constantly needed, but the point is clear: without verification from good experiments -- from whatever source -- the armchair theorizing that has typified generative phonology is mere game-playing. Quite a lot of fun, admittedly, but incapable of telling us very much about the speaker's mental processes, which is our set goal. The emphasis here on the need for an empirical foundation to phonetic and phonological work should not be taken as a denial or rejection of the need for speculative studies. It should be clear that both speculation and testing of one's speculations are needed; *neither* can be dispensed with.

Although I have some argument over the methodology adopted by generative phonology insofar as it purports to shed light on psychologically real entities and processes, I in no way deny the significant advances and contributions made by this school by way of describing and cataloguing the sound patterns of languages. Chomsky and Halle's *The Sound Pattern of English* (1968) presents many useful and hitherto unknown facts about regularities in English phonology. In addition much valuable work has been done in the generative phonology framework on such diverse languages as Japanese, French, Sanskrit, Akan, Russian, Spanish and many others. Distinctive feature notation itself, in its various forms, represents an advance over previous notations in that it allows broader, more general regularities in the sound pattern of a language to be stated most economically and in a way that is frequently quite revealing.

However speech research has now reached a stage where it needs and has produced a better way to represent the phonetic facts -- a way which does a much better job at giving an *explanation* of the phenomena involved. A few examples of sound changes or alternations due to known causes may make this clear.

1. According to Wang (1967), in an early stage of Chinese voiced and voiceless stops were differentiated but later the voicing distinction disappeared and was replaced by a tonal difference on the following vowel; low tone appearing where the consonant preceding had been voiced, and high tone when the preceding consonant had been voiceless. As Wang noted, this is easily explained by considering the mechanical properties of the vocal apparatus. Voicing during an occlusion tends to lower the fundamental frequency since the pressure build-up supraglottally tends to reduce the pressure drop across the glottis and consequently the air-flow rate through the glottis. Thus upon release of the voiced stop the fundamental frequency initially rises. In the case of the voiceless stops, however, particularly voiceless aspirated stops, i.e., those in which the vocal cords adduct some time after the release of the stop (of the order of 40-60 msec. or more), the vocal

cords are adducted in the presence of an initially high rate of air flow through the glottis and consequently start vibrating more rapidly (Sato 1958, Lehiste and Peterson 1961, Ladefoged 1967b, Chistovich 1969, Haggard 1969). Thus, upon the release of a voiceless aspirated stop the fundamental frequency will be falling. These tonal features were undoubtedly secondary auditory cues for the different lexical items at the early stage of Chinese but when the voicing distinction was lost became the primary distinctive feature differentiating the words.

2. For similar reasons, Punjabi now has replaced what was a breathy-voiced stop by a voiceless unaspirated stop plus a tonal contour. Interestingly, the tone is falling on the preceding vowel and rising on the following vowel, as we would expect, knowing that breathy-voiced phonation lowers fundamental frequency.

3. The reason so many of the world's languages have /i,a,u/ vowel systems if they are limited to only three vowels is because these vowels are maximally distinct from one another in their spectral properties (Wang 1968).

4. An overwhelming number of sound changes and sound alternations can be explained -- or at least rendered plausible -- if we consider the acoustic properties of the affected sounds before and after the change. For example, the substitution of /ʔ/ (glottal stop) for medial /t/ in many dialects of English, e.g., Cockney [bɒʔi] for "bottle", is no doubt due to the strong acoustic similarity between medial /t/ and /ʔ/, i.e., silence and minimal formant transitions on surrounding vowels. Similar reasons explain the change of /hɛnri/ to /ʔɛnri/.

Likewise, such considerations lend plausibility to such well-attested changes as:

$$\left.\begin{matrix}[\text{çi}]\\[\text{xi}]\end{matrix}\right\} \rightarrow [\text{hi}]$$

$$[\text{pʰ}] \rightarrow \left\{\begin{matrix}[\text{pf}]\\[\text{f}]\\[\text{h}]\end{matrix}\right.$$

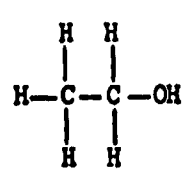
$$[\text{VN}] \rightarrow [\tilde{\text{V}}] \quad \text{(where V = vowel and N = nasal consonant)}$$

In fact it seems safe to say that most sound changes can take place only if they do not result in a drastic change in the acoustic identity of a word.

Many more examples could be offered, but this is not the place (however, see Schane 1969, for a more systematic analysis of diverse types of sound changes). My point is to illustrate particular instances of sound change or sound alternation which are rather easy to explain when represented in terms of the relevant physiology and

acoustics of speech. The account of such changes in terms of distinctive features may be adequate and even elegant in their ingenuity, but they are not explanatory. Apparently many of the inadequacies of this notation have dawned on the generative phonologists, too, and for this reason the marking convention was adopted (Chomsky and Halle 1968, p. 400 ff.). Specifically, it was recognized that the system failed to take into account the "intrinsic content" of the features. That is obviously true, as I have indicated above, but the proposed changes still go only a short way towards rectifying that deficiency. It strikes me that there is no possible way of modifying the distinctive feature notation, as long as it still represents sounds as a column of features and words as a matrix of such features, which will enable it to conveniently capture and explain the significant aspects of the sound changes and sound alternations pointed out above. Instead it is necessary to represent these processes by mathematical or hardware analogues of the vocal apparatus, analogues which allow us to incorporate all the known mechanics and constraints of speech physiology and speech acoustics. Such models have been constructed and used profitably by Fant (1958), Öhman and Lindqvist (1966), Flanagan and Landgraf (1967 and 1969), Öhman (1967a), Stevens (1968), Umeda and Teranishi (1966), Lindblom (1968a) and many others.

The importance of a notation system in any scientific discipline, i.e., the method whereby the researcher can represent and work over the accumulated data in simplified form, cannot be overemphasized. The particular method chosen may stimulate further advances if it is optimal or inhibit further progress if it is poorly adapted to the purposes and needs of the field. The tremendous effect on mathematics of the adoption of the Hindu number system to replace the Roman number system is well known. Similarly, in chemistry, the usual two-dimensional letter-and-line symbols for molecular structure, e.g., that for alcohol,



proved to be incapable of conveniently representing important three-dimensional structural aspects of the molecule, e.g., the distance and angle of the bond between one atom and another. Precisely machined balls and sticks are an important part of modern chemical notation as well as being convenient vehicles for research.

Any attempt to improve the notational system for representing phonological or phonetic facts must be encouraged. But it appears to be unprofitable to burden any notational system, and in particular the distinctive feature system, with the task of representing and explaining all the relevant facts of historical sound change, perceptual and physiological aspects of sound alternation in addition to aspects of the acquisition and development of sound systems in children.

Further it is carrying things a bit far to reject without discussion, as Chomsky and Halle do (1968, p. 327-29), a perfectly good, experimentally-founded analysis of voicing distinctions in the world's languages, namely that of Lisker and Abramson (1964), in favor of another system involving such largely uninvestigated features as "tenseness", "heightened subglottal pressure", and "glottal constriction". (See the reply by Lisker and Abramson 1968.)

Recalling again Stevens' evaluation of the field of speech research, it is impossible to tell when we will have our own analogue to the periodic table of elements, but one thing is clear: we won't have it until we build up a structure of reliable facts and models comparable to the one chemistry had before the periodic table was proposed. For the present, though, it seems unavoidable that we shall profit from the current trend to investigate the mental processes and brain mechanisms underlying observed speech behavior. Our task, then, is to try "to see how the human mind functions" (Lieberman 1967a, p. 2). There is some hope that the dividends of this approach will extend beyond helping us to have a better understanding of how speech is controlled and produced, and how it is subject to change. Lashley (1951) reminded us of Fournie's (1687) comment that "Speech is the only window through which the physiologist can view the cerebral life." Whether it is the only "window" is debatable, but that speech is excellent subject matter to reveal the workings of the brain is a reasonable hope. It may be that by studying the mechanics underlying speech we may be able to find out how voluntary motor behavior is stored latently, selected, activated and controlled -- a problem that neurophysiology has attacked for many years using, for the most part, animals. The movements of speech have the advantage over most studies of movement in animals -- whether decerebrate or not -- in that they are voluntary, executed under control of the cerebrum, are repeatable and quantifiable. Perhaps the best example of how phonetics can exploit these qualities to its advantage is the work of the group in the Pavlov Institute of Physiology (Chistovich, et al. 1965, Ventsov, et al. 1966).

But, again, if we can take any lesson from the other "successful" sciences, our understanding of the speech process will have to proceed bit-by-bit, even if we are lucky enough to have an occasional breakthrough. We cannot just build theories in the sky and let the testing of them wait until later.

In the present work Chapter Two attempts to shed light on the mechanisms speakers use to control the fundamental frequency of phonation in speech. Instrumental investigations of up to five subjects are reported which lead to the re-affirmation of the traditional view that fundamental frequency is controlled almost exclusively by the laryngeal muscles. Claims by Lieberman (1966, 1967a, 1967b, 1968a, 1968b)

challenging this view are examined in detail, and an attempt is made to counter them. Chapter Three considers arguments, evidence, and experimental techniques relevant to discovering certain possible brain mechanisms underlying observed speech behavior. In particular, two issues are covered: the possible role of feedback in speech and how the timing of gestures in speech is controlled.

Chapter 2

The previous chapter expressed the view that the task of phonetics is to discover the stored units of speech, explain how they are manifested and controlled as a pattern of muscular contractions, how they are tailored for optimum communication, and how they are perceived and identified by the listener. Viewed in this way, phonetics is but one specialized branch of the behavioral sciences. As in any other science this complex of tasks requires (a) constructing a model of the system we wish to understand and (b) seeing if the hoped-for correspondences between the output of our model and the empirical observations of the real world can be demonstrated.

Although relatively inaccessible, the action of the larynx and one aspect of its acoustic output, fundamental frequency (F_o) or "pitch," has, perhaps, received more attention and study and has more often been modelled than have any of the other physiological mechanisms in the speech apparatus. We know much more about the workings of the larynx and the way in which pitch varies in speech than we do about seemingly more accessible articulators such as the lips and the lower jaw. This is due no doubt to the fact that pitch is relatively easily transduced into graphical form, being readily extractable from the acoustic signal of phonation, whereas lip movement is not, and due to the fact that a separate medical discipline, laryngology, exists which gives its special professional attention to the workings of the larynx. But the existence of a separate medical discipline has unfortunately meant a neglect of physiological studies of laryngeal activity during speech. The usual subject of physiological investigations have been how the larynx executes variatons in pitch, intensity, or register during steady-state phonation and similar non-linguistic actions.

However we should not overlook two other important factors which help to explain the apparent lack of attention to speech. First, the best methods of studying the physiology of the larynx necessitate either that the larynx be removed from the body, or if not removed, that the host be later sacrificed (thus necessitating animal subjects), or that the vocal tract itself be occupied by the instruments used. These techniques rule out study of normal speech. Second, it has been assumed, implicitly and explicitly (Arnold 1961) that what was found to be true of singing or non-linguistic use of the vocal mechanism, should apply as well to speech. This is a reasonable assumption, but until demonstrated experimentally it would have to remain merely an assumption. As

we shall see below, Lieberman (1966, 1967a, 1967b, 1968a, 1968b) raises doubts that this assumption is true, suggesting that the way F_0 is regulated in speech is different from the way it is controlled in singing or other non-linguistic phonation.

The main issue of manner of control of F_0 in speech is, then, one of finding out if the principles which have repeatedly been demonstrated to operate during non-linguistic phonation also hold in speech.

Such physical parameters as vocal cord tension, the mass free to vibrate, losses due to collision of the surfaces of the two vocal cords, air flow through the glottis, etc., all determine the rate of vocal cord vibration. These are not easy to obtain in quantifiable form in the living, speaking subject although they can be quantified in excised larynges. But fortunately some secondary, indirect indications of the vocal cord tension, the mass free to vibrate, etc., can be sampled in the living subject, e.g., the length of the vocal cords, the vertical displacement of the larynx and the electrical activity in the laryngeal muscles.

STUDIES ON EXCISED LARYNGES*

Ferrein (1741), in experiments on excised larynges, established that increasing the tension of the vocal cords increases the pitch of the sound produced (provided, of course, there is a current of air sufficient to drive the vocal cords). His results were replicated by Müller (1843) who showed in addition that pitch could also be increased by augmenting the force or pressure of the air driving the vocal cords, although it is clear from his quantitative data that changes in tension were capable of varying pitch over a much wider range than were changes in air pressure. These results have been obtained repeatedly by other investigators (Oertel 1895, Réthi 1896, van den Berg and Tan 1959, Anthony 1968). One obvious difficulty in working with excised larynges is, as van den Berg and Tan have remarked, the fact that the muscles cannot contract or maintain their usual tonus. Some simple movements can be duplicated by the pull of strings appropriately tied to the muscles or cartilages, however the action of the vocalis muscle cannot be adequately copied. This doesn't invalidate all the results obtained from excised human larynges, though, because similar results have been obtained with animal larynges *in vivo* (Isshiki 1959, Hast 1961, Koyama, Kawasaki and Ogura 1969). The studies on animal larynges have been particularly useful in elucidating some of the inter-relatedness of the parameters determining vocal pitch. For example, with a constant air flow an increase in the stimulation of the adductor muscles produces an increase in subglottal pressure, intensity, and pitch. Likewise, of course, a constant stimulation rate plus an

* Illustrations of the anatomical structures referred to may be found in the appendix.

increase in air flow produces increases in the same parameters. Further detailed findings from excised human larynges will be mentioned below.

Electromyographic and Clinical Studies of Laryngeal Activity

Electromyographic studies of the cricothyroid show that it exhibits increasing activity during rising pitch (Katsuki 1950, Faaborg-Andersen 1957 and 1965, Sawashima, Sato, Funasaka, and Totsuka 1958, Zenker and Zenker 1960, Arnold 1961, Öhman, Mårtensson, Leanderson, and Persson 1967, Perkins and Yanagihara 1968). Katsuki found some cases of increased cricothyroid activity during low pitch. Zenker and Zenker characterize the pattern of activity of the cricothyroid as being low for the middle range of pitch but increasing both for low pitch and high pitch. Perkins and Yanagihara found this in one case out of several. Wenrick (1931) applied electrical stimulation externally in the vicinity of the cricothyroid muscle, and, noting that the pitch of phonation was lowered, concluded that a contraction of the cricothyroid lowers pitch.

Electromyographic studies of the intrinsic laryngeal muscles generally find an increase in activity in all the adductor muscles during rising pitch (Faaborg-Andersen 1957 and 1965, Sawashima, et al. 1958, Öhman et al. 1967).

Although the extrinsic laryngeal muscles have been extensively investigated during phonation (Michel 1954, Zenker and Zenker 1960, Faaborg-Andersen and Sonninen 1960, Sonninen 1956, Arnold 1961, Kimura 1961, Faaborg-Andersen and Vennard 1964, Hirano et al. 1967) there is still some doubt as to the action of most of these muscles. The sternohyoid shows least activity in the middle of subjects' vocal ranges and increasing activity for low or extremely high pitch (usually above the range used in speech) (Hirano, et al. 1967). The function of an increase for high pitch is not clear; the increase for low pitch is probably related to the common observations that the vertical movements of the larynx correlate well with the variations in the pitch of phonation (Garcia 1840) and that the sternohyoid pulls the larynx downward. The reason for this relation between larynx height and F_o is not completely understood either. This phenomenon is discussed further below.

Hirose (1969, personal communication) reports a case in which a patient who had lost the use of his cricothyroid muscles, but not the use of his other intrinsic laryngeal muscles (by an accidental bilateral section of his external superior laryngeal nerves during a thyroidectomy), could still manage some variations in pitch but complained he had difficulty in executing high pitch in speech and couldn't achieve falsetto in singing at all. Similar cases have been reported

and reviewed by Luchsinger and Arnold (1965). There is thus much redundancy in the muscular system regulating pitch such that if one or two muscles are lost, others can take over and pitch regulation is not thereby completely lost, although it may be drastically restricted in range. However, that *some* laryngeal muscles must remain in working order for pitch regulation is suggested by some cases of progressive bulbar palsy in which the motor cranial nerves, including those supplying the laryngeal muscles, degenerate gradually leaving a patient who may be capable of phonation but without the ability to vary pitch (Critchley and Kubik 1925).

An estimate of the effectiveness of the respiratory muscles in regulating pitch by varying the pressure or flow of the air through the glottis may be made by considering the ability of patients with respiratory paralysis to vary the pitch of phonation. Ladefoged (personal communication) has studied both paraplegics with paralysis of all the normal expiratory muscles (see Draper, Ladefoged and Whitteridge, 1960) and poliomyelitis victims who found themselves in an "iron lung" which took over the function of the paralyzed respiratory muscles. He reports that neither group had noticeable difficulties in pitch control. Severe paraplegics relying principally on the elastic recoil of the lungs after a diaphragmatic inspiration could pronounce short sentences perfectly; and poliomyelitis victims had no great difficulty in speaking naturally provided they learned to speak only during the expiratory phase of the machine's pumping cycle. The "iron lung," of course, does not provide any short-term variations in the respiratory force and thus whatever pitch variations these patients executed had to be due to the action of their laryngeal muscles alone.

(Additional testimony on the naturalness of speech produced in an iron lung has been offered by Peterson (1958) and by those who listened to his demonstration (Wang 1962, Martin 1956).

Radiographic and Photographic Studies of Laryngeal Activity

Such quantifiable measures as changes in the length of the vocal cords, in the thickness of the vocal cords, and in the size of the laryngeal ventricles (ventricles of Morgagni) are indirect indications of vocal cord tension and the mass free to vibrate and can be obtained fairly simply from sagittal and frontal X-rays of the larynx, as well as from laminograms and from photographs of the vocal cords via indirect laryngoscopy (French 1884, Moeller and Fischer 1904, Farnsworth 1940, Griesman 1943, Mitchinson and Yoffey 1948, Sonninen 1954 and 1956, Hollien and Curtis 1960, Keenan and Barret 1962, Fink 1962, Ohala and Vanderslice 1965, Ardran and Wulstan 1967). Such studies have provided repeated confirmation of the general principle that pitch is varied by changing the length and consequently the tension of the vocal cords, and by changing the mass that is free to vibrate.

The good correlation between the vertical displacement of the larynx-hyoid bone complex and the pitch of phonation in speech and singing has long been known (Garcia 1840) and is easily noticed in speakers with a

prominent "Adam's apple." Some authors doubt that the vertical movement of the larynx with pitch is necessary or desirable for a "trained" singer (Luchsinger and Arnold 1965) but that it commonly occurs in the pitch range used in speech or in speech itself has been demonstrated and quantified by Inoue (1931), Griesman, Keenan and Barrett, Amenomori (1960 and 1961), Sonninen (1956 and 1968), and Vanderslice (1967), and is evident in many other authors' published data (Arnold 1961). There is little doubt that there are consistent patterns of muscular action in the extrinsic laryngeal muscles and the supra-hyoid muscles in the regulation of pitch in speech. This in itself however does not tell us the relative contribution of these muscles to observed pitch changes. Many muscles attach to the larynx-hyoid-bone complex and undoubtedly influence its position. All of these forces have not yet been identified or factored out. Roughly speaking, though, it appears that the height of the larynx varies directly with the log of the pitch of phonation in the lower end of a speaker's pitch range, i.e., the range used in speech. The reasons for this is not known exactly but some interesting speculations have been presented by Sonninen. Upward and downward movement may contribute somehow to lengthening or shortening, respectively, of the vocal cords, or it may assist in varying the mass that is free to vibrate. Quite clearly this is not the sole mechanism of pitch regulation since patients having their sternohyoids, omohyoids and/or sternothyroids severed can still execute pitch variations in speech and singing, although perhaps over a more restricted range than before the loss of these muscles (Sokolowsky 1943, Sonninen 1956 and 1968).

Aerodynamic Studies of Laryngeal Activity

It is not difficult to record subglottal air pressure in speech. This can be done either by indirect means, e.g., by recording the esophageal pressure (van den Berg 1956, Ladefoged 1962 and 1967a, Strenger 1959, Lieberman 1967a and 1968c, Vanderslice 1967) or by using a body plethysmograph (Kozhevnikov, et al. 1966) or by the direct method of inserting a catheter or needle into the trachea (greatly facilitated if the subject already has a hole in his trachea from a previous operation) and connecting the catheter to a pressure transducer of some sort (Rousselot 1924, Smith 1944, Stetson 1928, van den Berg 1956, Fischer-Jorgensen and Hansen 1959, Isshiki 1959, Ladefoged 1960, Ohman and Lindqvist 1966a and 1966b). It is now also possible to insert miniature pressure transducers directly into the trachea (Koike and Perkins 1968). With the exception of Stetson's claims that a separate breath pulse or subglottal pressure peak is present for each syllable (a claim which was falsified by Ladefoged, Draper, and Whitteridge 1958, and more recently by Lieberman, Griffiths, Mead, and Knudson 1967) these investigations yield much the same picture of subglottal pressure in speech. It necessarily becomes positive during expiratory phonation, usually rising rapidly before and falling rapidly after phonation, but maintaining a fairly constant

level in between. Momentary increases in subglottal pressure are noticeable on emphasized ("stressed" or "accented") syllables and have been shown by electromyography to be due to some extent to contractions of the expiratory muscles (Ladefoged 1962). Some momentary variations in subglottal pressure, however, can be attributed to short term variations in the glottal resistance which is regulated by the laryngeal muscles (Ladefoged 1963, Broad 1968). Some variations in subglottal pressure, particularly short term variations, are thus a function both of the respiratory apparatus (active expiratory effort plus the passive effect of the elastic recoil force of the lungs and thorax) as well as the mean glottal resistance.

Increased air flow (or subglottal pressure) has been shown to increase pitch in the living subject in the investigations of van den Berg (1957), Isshiki (1959), Ladefoged (1963) and Öhman and Lindqvist (1966a). These studies all showed that a push on the chest or abdomen of a subject attempting to phonate at a steady pitch level causes the pitch to rise involuntarily and immediately. This effect is certainly familiar to anyone who has attempted to sing or carry on a conversation while driving over a bumpy road. If one assumes that the laryngeal adjustment is constant in these cases, the measured value of $\Delta F_0/\Delta P_s$ (with F_0 in Hertz and P_s in centimeters aqueous, i.e., centimeters of water) should be a fairly accurate quantification of the effect of subglottal pressure on F_0 for a given adjustment of the larynx. Table I gives the values of this expression from the experiments of Isshiki (1959), Ladefoged (1963), Öhman and Lindqvist (1966a), van den Berg and Tan (1959), Furukawa (1967), and Anthony (1968), the last three working with excised human larynges.*

As mentioned above, excised larynges lack the usual muscular tone in the vocalis and other parts of the thyroarytenoid muscle. Since contraction of the vocalis can significantly affect the glottal resistance which in turn affects the rate at which a given level of P_s can affect F_0, values of $\Delta F_0/\Delta P_s$ obtained in such studies may be too high.

On the basis of such a "calibration" of the effect of variations in subglottal air pressure, only Ladefoged (1963) and Öhman and Lindqvist (1966a) actually made any specific claims about the relative contribution of the entire laryngeal musculature and the subglottal respiratory musculature towards effecting the observed pitch changes

---

* Values of $\Delta F_0/\Delta P_s$ are also available for non-human larynges, e.g., Hast (1961) working on dogs' larynges *in vivo* obtained values in the range 2.3 to 2.9 Hz/cm. aq. However, these may not be directly relevant quantitatively to studies of human larynges.

Cf. also: Negus, V.E. (1949), p.142.

Table I. Calibrations of $\Delta F_o/\Delta P_s$ (in Hz/cm. aq.)

Source	Normal Voice	Falsetto	Conditions
Isshiki (1959)	3.3	--	1 sample from 1 subject, male; pressure obtained from tracheal needle
Ladefoged* (1963)	5	---	1 subject, male; pressure estimated from esophageal balloon
Öhman and Lindqvist** (1966a)	2.5	---	1 subject, male; pressure obtained from tracheal needle
van den Berg and Tan (1959)	5-12	17-20	excised male larynx representative of some 30 other larynges
Furukawa (1967)	0-16	--	excised larynx
Anthony (1968)	6-8	---	excised female larynx

in speech. Using the value of 5 Hz/cm. aq. Ladefoged factored out those pitch variations attributable solely to changes in the pressure

* Ladefoged reported the value of 1/2 octave/7.5 cm.aq., but his figure suggests that it was more like 1/2 octave/9 cm. aq., which over the pitch range covered, equals 5 Hz/cm. aq. In addition Ladefoged and McKinney (1963) in an earlier article had reported sampling the subglottal pressure and F_o of a subject producing monosyllabic words at various intensities with unrestrained pitch. They noticed a good correlation between the peak pressure and peak F_o of each word and the value of $\Delta F_o/\Delta P_s$ was very close to that found by Ladefoged (1963), namely, 7.5 Hz/cm. aq., even though there was no attempt to insure an unvarying laryngeal adjustment in this case. They noted that "it is interesting to speculate on the extent to which changes in intonation in English might be due to variations in subglottal pressure rather than variations in the tension of the vocal cords." (p. 458)

** Öhman and Lindqvist reported that the period of the glottal vibration diminishes by .16 msec./cm. aq., which works out to an increase in pitch of approximately 2.5 Hz/cm. aq. in the pitch range studied.

drop across the glottis (especially owing to supraglottal pressure increases during consonantal articulations) and thus derived another curve labelled "vocal cord tension," defined as "the sum of the physiological factors affecting the adjustment of the larynx" (p. 118). By far the greater part of the gross variations in F_o were attributable to this parameter.

Öhman and Lindqvist, having determined the value of 2.5 Hz/cm. aq, experimentally, similarly factored out those pitch variations that could be attributed to the changes in the pressure drop across the glottis and derived curves representing the action of the laryngeal musculature. Again, most of the observed pitch variations were assigned to the action of the larynx. The pressure rises on stressed syllables were said to be capable of causing at most only about 10% of the observed pitch rise. They noted as well that

> ...the ΔP change [i.e., the change in the pressure drop across the glottis] which is due to stress is always much smaller than that due to stop consonants for instance, and ... the f_o changes during the stressed syllables do not correlate well with the stress-induced ΔP changes either in phase or in amplitude. [p. 4]

Relevant to this issue are the studies of Flanagan and Landgraf (1967 and 1969) and Flanagan (1968), who, on the basis of calculations from a mathematical vocal cord model suggested that acoustic coupling of the vocal tract and the vocal cords may become more important with constricted vocal tract shapes such as that for /u/ and that therefore certain values of subglottal pressure would have a greater effect on F_o if the vocal tract configuration were constricted than if it were relatively open. Isshiki's, Ladefoged's and Öhman and Lindqvist's calibrations were performed while their subjects phonated open vowels and thus would not show this coupling effect if it does exist. In applying such values of $\Delta F_o/\Delta P_s$ to real speech situations it must be remembered that this value varies to some extent with the mean glottal resistance. Glottal resistance may undergo considerable short-term (100-200 msec.) variation in running speech, as is evident by the momentary drops in pressure during [h]'s. Also Isshiki found that the minimum subglottal pressure required to maintain phonation varied directly with the pitch, which suggests that the mean glottal resistance also varies with pitch. This was confirmed experimentally by Yanagihara and von Leden (1966).

Vanderslice (1967) made a qualitative estimate of the relative contribution of the subglottal air pressure and the extrinsic laryngeal muscles. He recorded the esophageal pressure, larynx height, and F_o (although not simultaneously in the case of the first two parameters) and found the larynx height (a measure of the activity of extrinsic laryngeal muscles which affect pitch adjustment -- see above) to be in better correlation with the F_o than the pressure was.

He concluded that "intonational pitch control is primarily vested in the larynx."

The conventional view of pitch control was challenged however by Lieberman (1966, 1967a, 1967b, 1968a, 1968b). He sampled and recorded the subglottal pressure and the F_o for three speakers' productions of several sentences with various intonation contours. On some sentences (statements and WH-questions, i.e., those having a terminal fall in pitch) he noticed a good correlation between the pressure and the F_o; on others (yes-no questions, or those having a terminal pitch rise) the correlation was not very good, especially at the end where the pressure fell but the pitch rose. He concluded that in the first class of sentences the correlation was of a causal nature, i.e., the subglottal pressure fully determined the pitch variations and the laryngeal contribution was assumed to be zero or negligible. When the correlation was poor, as it was for the terminal portions in the other class of sentences, it was allowed that the laryngeal musculature was active in influencing the pitch.

Lieberman employed the familiar term "breath-group" to refer to the complex coordinated physiological activity which goes on in between successive inspirations in speech and which initiates and supports phonation and produces the observed variations in intensity and pitch (or at least those which cannot be attributed to air pressure transients due to supra-glottal articulations). Borrowing the terms and concepts of "marked" versus "unmarked" from Chomskyan and/or Praguian linguistics he referred to the first set of sentences as "unmarked breath-groups" (symbolized [-BG]) since they are alleged to involve the least expenditure of effort and are supposedly simpler, intonationally, than the other sentences, which thus become the "marked breath-groups." That is, Lieberman suggests that if the speaker doesn't do anything extra with his laryngeal muscles he winds up with an unmarked breath-group (and the pitch falls at the end when the subglottal pressure does). If he does perform an extra bit of work with his larynx at the end of the sentence, he gets a "marked breath-group" (and the pitch rises in spite of the fall in pressure). Further, Lieberman's data reveal the well-known momentary increase in subglottal pressure and the concomitant increase in F_o when a syllable is emphasized or receives prominence (symbolized [$+P_s$]). This pitch rise, like the pitch fall of [-BG] he attributes solely to the effects of the subglottal pressure.

Obviously any claim that in some environments F_o is determined by subglottal pressure and only negligibly by laryngeal adjustment must rest on the more fundamental claim: that the observed variations in subglottal pressure are themselves not determined by laryngeal adjustment. Lieberman in fact does not address himself to this issue nor does he attempt to prove or justify it logically or experimentally. It is however the logical basis on which his other claims rest and must be considered as part of his hypothesis on the physiological basis of

intonation.

Lieberman made no attempt to calibrate the effect of subglottal pressure on F_o, but having merely assumed that the laryngeal muscles were inactive with respect to pitch in the unmarked sentences, he plotted (in Lieberman 1967a) the F_o against the corresponding subglottal pressure at 30-odd points and arrived at values for $\Delta F_o/\Delta P_s$ between 16 and 22 Hz/cm. aq.,* which are from 2 to 8 times larger than the values obtained for normal voice by the authors cited above, with the exception of the values calculated by Flanagan and Landgraf (1967) and Flanagan (1968) from their model of the vocal cords. Lieberman recognized this disagreement between his values for the rate at which subglottal pressure could affect F_o and those published by others who had experimented with actual larynges. He suggested that this could be attributed to a basic difference in the manner of control of F_o in speech and in singing (the "calibration" such as Ladefoged and Öhman and Lindqvist performed having not been obtained during speech as such).

In some of Lieberman's records the F_o does not follow the subglottal pressure in the way predicted. To account for these discrepancies he suggests that there are "archetypal"** physiological correlates

---

* For some unexplained reason, in the summary on p. 106, the range of values for $\Delta F_o/\Delta P_s$ is reported as being 17-20 Hz/cm. aq. This is, though, the least of the inconsistencies between the main body of the book and the summary. On pp. 95-96 it is admitted that the data points for these graphs

> ... were measured ... at points where we *assumed* that the tension of the laryngeal muscles was unchanged. [italics mine]

But in the summary on pp. 106-107 this assumption is suddenly *proved* by the data it was based on:

> These [graphs] *showed that* the tension of the laryngeal muscles was relatively constant during the nonterminal portions of the breath-group. [italics mine]

** The use of the term "archetypal" is particularly unfortunate. It might be thought to denote merely something like "basic," "classic" (in the colloquial sense), "normal," or "statistically frequent." However, it connotes much more, such as "innately determined," "primordial," or "primeval." This would perhaps be appropriate when applied to the "unmarked breath-group," the characteristics of which Lieberman does claim are innately determined, however this term is

of the basic phonological units (or "features"*), the "breath-group" ([BG]) and "prominence" ([P_s]), and that the speaker and hearer encode and decode speech with reference to archetypal or more basic forms of these abstract units; however the speaker may also employ "alternate articulatory maneuvers" which are physiologically different from the archetypal forms but which produce the same acoustic output. Communication is not thereby impeded, the argument runs, because the acoustic shape of the unit is all that ultimately matters for successful communication. Thus the "archetypal" form of an unmarked breath-group would require that the subglottal pressure fall at the end and that the F_o fall simultaneously. If, however it is observed that in a particular case the subglottal pressure does not fall at the end but the F_o falls anyway, (as happens in figure 4.25, p. 86, Lieberman 1967a), this is explained as the speaker employing an "alternate articulatory maneuver" (in such a case, changing the tension in his vocal cords) which still results in the intended acoustic correlate, the terminal falling pitch. In this way Lieberman neatly accounts for all the data which fit his theory and seemingly for those which don't, as well.

Other evidence is adduced by Lieberman to show that his posited features [+P_s] and [$\pm$BG] are psychologically real and that [-BG] is innate and thus universal to all human languages. This extra evidence included studies on newborn infant cry, a psycholinguistic experiment of Hadding-Koch and Studdert-Kennedy (1964), and evidence on the intonational patterns found in many diverse languages.

Although Lieberman has offered his theory of intonation as being

---

also applied to the "marked breath-group," i.e., the increase in the tension of the laryngeal muscles at the end of a question, and to the "marked" state of the feature [P_s], i.e., the momentary increase in the subglottal pressure. There is no evidence offered that the characteristics and the manner of using the latter units are innately determined.

* The term "feature" as used in most current works espousing generative phonology, means something more like "unit" or "primitive entity" in forming contrasts, rather than the usual dictionary meaning of "feature" which is "... a distinct, or outstanding part, quality, or characteristic of something ..."

"verified" Lieberman (1968a, p. 157) and although his theory is being confidently accepted in some current literature, his claims are obviously far reaching enough to merit close examination and attempts at replication. The claim that the laryngeal muscles are negligibly involved in producing the observed pitch variations in speech (except for the terminal rise in yes-no questions) is exactly opposite to the traditional view maintained by such phoneticians as Sweet (1877, p. 93) and Fry (1964, p. 217-218), the view which has been lent support by the experiments of Ladefoged, Öhman and Lindqvist, and Vanderslice.

It was with the intention of shedding some light on the issue of the extent of participation of the laryngeal muscles in regulating pitch in speech that the studies to be reported below were undertaken.

* * * *

The investigations reported here were conducted in collaboration with Dr. Minoru Hirano. The purpose of these studies was quite simply to assess the relative contribution of the larynx in producing the observed variations in the F_o of phonation during speech. There were three separate studies.

STUDY ONE: EMG FROM LARYNGEAL MUSCLES OF FIVE SUBJECTS

In the first study electromyographic recordings from various laryngeal muscles were obtained in five subjects during the production of various English sentences incorporating a variety of prosodic patterns. Table II indicates which muscles were recorded from in which subject.

Table II. Schedule of Muscles Studied in Five Subjects

	crico-thryoid	lateral crico-arytenoid	vocalis	inter-arytenoid	posterior crico-arytenoid	sterno-hyoid
JO	X	X	X	X	X	X
GA	X	X				X
DB	X	X				X
WV	X	X	X	X		X
LC	X	X				X

The extent to which any given muscle was investigated in all these subjects is also an indication of the confidence with which conclusions are stated about the behavior of that muscle in regulating F_o. All subjects were adult speakers of Standard American English and all but subject LC were male. Subjects JO, GA, DB, and WV had all had phonetic training and subjects WV and LC were professional singers. All had normal larynges and were free of any speech defects.

Although many different sentences were spoken by each subject, the following are the items used by all five subjects:

(1) Bev bombed Bob. (As a possible answer to "What happened?")

(2) Bev bombed Bob. (As a possible answer to "Who bombed Bob?")

(3) Bev bombed Bob. (As a possible answer to "What did Bev do to Bob?")

(4) Bev bombed Bob. (As a possible answer to "Who did Bev bomb?")

(5) Bev bombed Bob? (Echo question to sentence #1)

(6) Did Bev bomb Bob?

(7) Did Bev bomb Bob?

(8) Bob. (Elliptical reply to " Who did Bev bomb?")

(9) Bob? (Echo question to sentence #8)

Subjects JO, GA, and DB also spoke the sentence "Joe ate his soup," with similar variations in accent and in question versus statement format.

The "Bev bombed Bob" series was chosen since only sentences which minimized the involvement of the laryngeal muscles for "segmental" gestures could hope to reveal those muscles' participation in purely prosodic gestures. We attempted to eliminate phrases including (a) word-initial vowels, since they are often initiated with a glottal stop and would require action from the vocalis, (b) voiceless consonants and /h/ since they both involve activity of the lateral cricoarytenoid and the interarytenoid, and (c) lingual consonants, since they seemed to involve noticeable activation of the sternohyoid. In fact it turned out to be impossible to completely eliminate segmental involvement of the sternohyoid since it is active for tongue retraction and jaw opening, both of which were present in our test sentences. Still, however, its participation in prosodic gestures could be evaluated by comparing sentences which were identical except for variations in pitch.

No special instructions were given to the subjects as to how to read these sentences. There were, as a result, some minor differences

in the prosodic patterns used, especially in the various question forms. However this in no way interfered with the main purpose of the study which was simply to find out if the laryngeal muscles regulated the various pitch modulations in speech or not.

In the course of this investigation a new technique was developed for obtaining electromyographic records from the intrinsic laryngeal muscles. A full description of this technique can be found in Hirano and Ohala (1969). Briefly, this technique involves a combination of the hooked-wire electrodes similar to those described by Basamajian and Stecko (1962) and the method of transcutaneous insertion into the intrinsic laryngeal muscles as developed by Hiroto, Hirano, Toyozumi, and Shin (1962) and as used by these authors for the first electromyographic study of the laryngeal muscles in speech (Hiroto, et al. 1967). Transcutaneous insertion, i.e., inserting the electrodes through the skin and other tissues of the neck (instead of passing the leads through the mouth and pharynx) has the advantage of permitting the subject to articulate normally. It has the evident disadvantage that, with the exception of the interarytenoid muscles and the vocalis muscle (where the leads cross through the laryngeal cavity) one cannot visually verify that the electrodes are in the desired muscle. However correct placement is possible by means of palpation, knowledge of the anatomical environment of the target muscle, and, finally, having the subject execute gestures known to involve certain muscles and not others (e.g., coughing, quiet breathing, singing musical scales, etc.), and, comparing the obtained EMG activity with previous records from the same and other muscles. Figures 2, 3, and 4 show the paths of insertion for some of the laryngeal muscles.

The articles cited above do not explain the method of inserting the electrodes into the sternohyoid. This muscle can be palpated on the neck and stands out particularly well when the subject strains against an upward force on the underside of his chin. The electrodes are inserted at the level of the middle of the thyroid cartilage and as close to the midline as possible and, as with the other muscles, as parallel to the muscle fibers as possible. In this area the only danger of contamination is from the omohyoid, and possibly the thyrohyoid. These can be easily avoided if the insertion is kept close to the midline. Two thin wire electrodes with hooked ends are inserted into the muscle with an ordinary hypodermic needle which is then withdrawn, leaving the hooked wires in place. The hooks in the wire are strong enough to prevent electrode slippage during ordinary speech movements, but are not so strong that they won't unbend and slip out painlessly at the end of the session when they are pulled firmly. These electrodes, once embedded in the desired muscle, cause the subject little or no discomfort, so that up to two or three hours of electromyographic recording can be obtained. Further, the bi-polar wire electrodes provide greater localization of the area from which

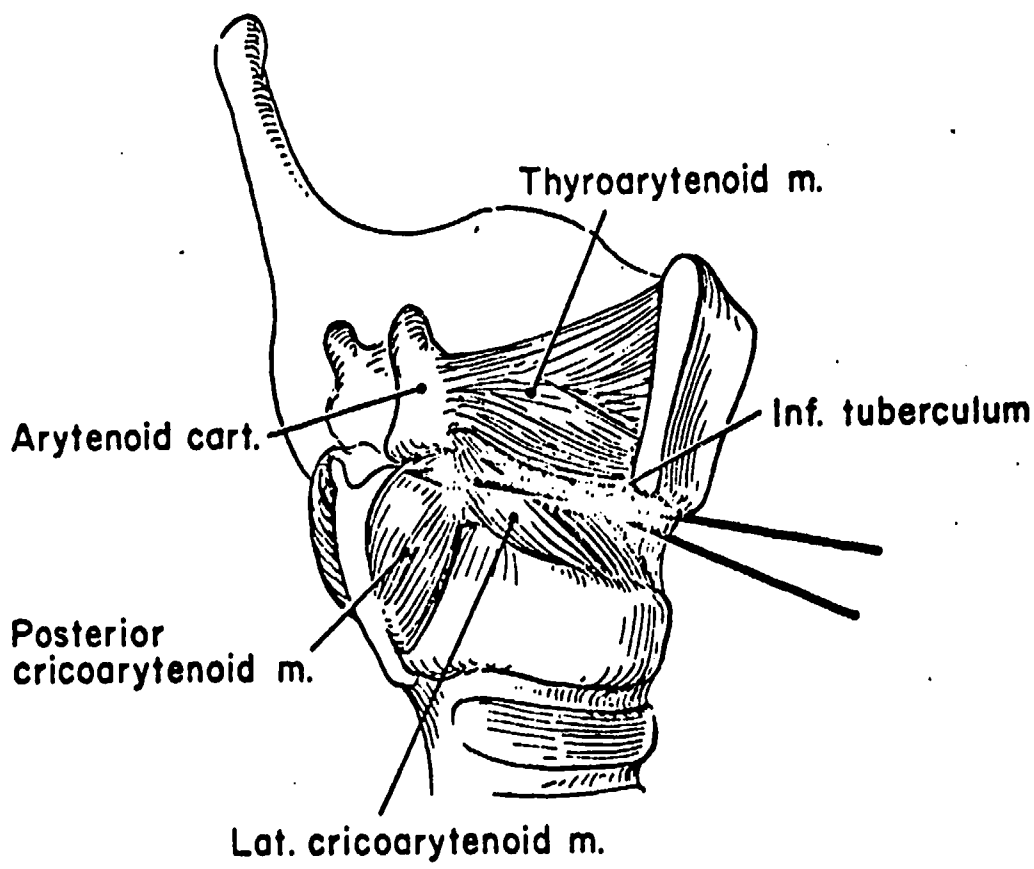


Figure 2. Direction of insertion of needle into the lateral cricoarytenoid and posterior cricoarytenoid muscles. (From Hirano and Ohala, 1969)

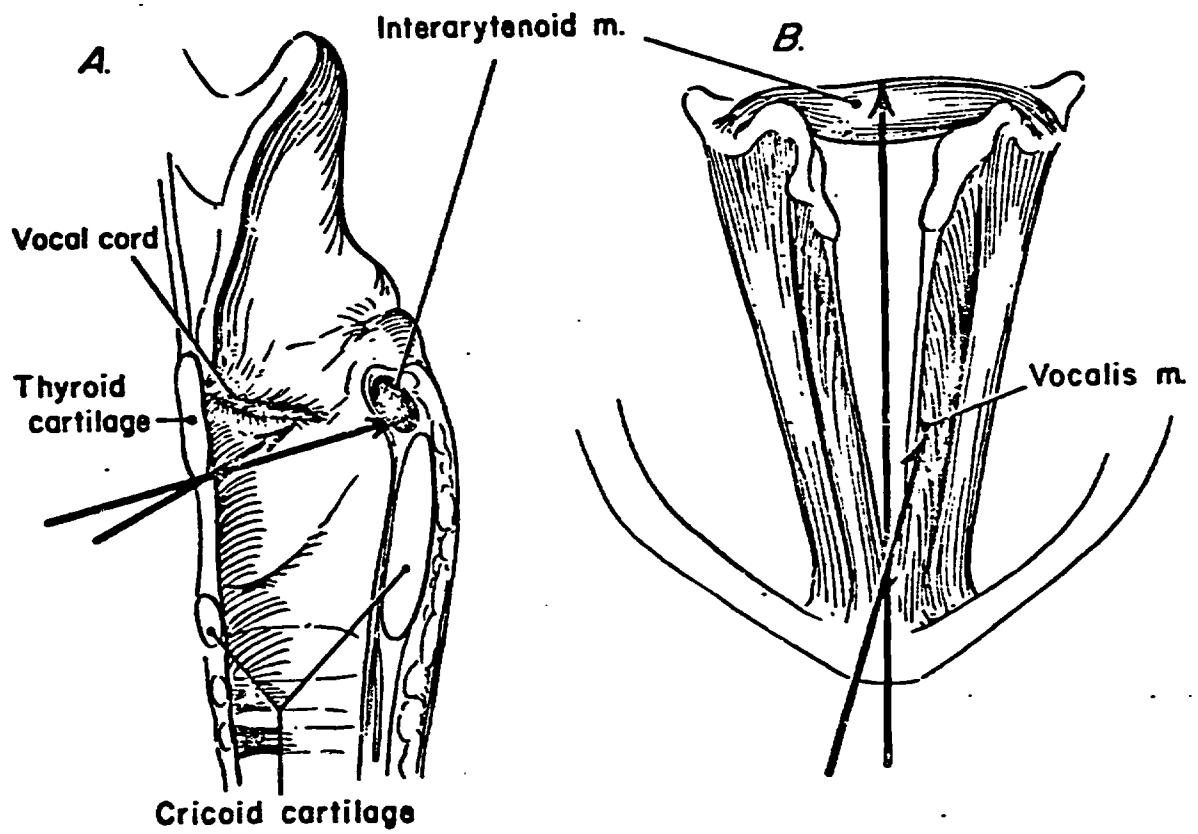


Figure 3. Directions of insertion for putting needles into the vocalis and interarytenoid muscles. (From Hirano and Ohala, 1969)

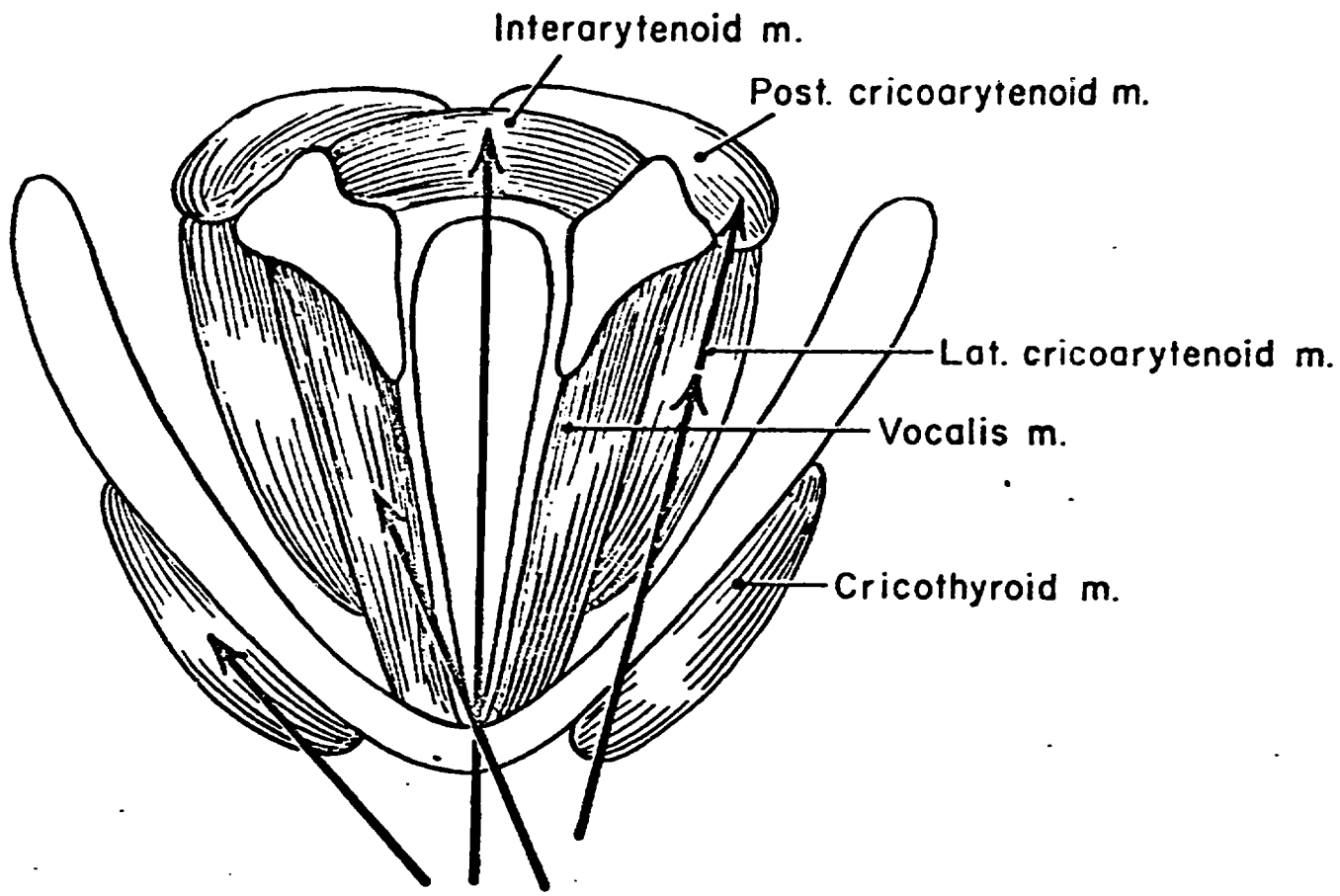


Figure 4. Schematic diagram, from above, showing directions of insertion for placing needles into all the intrinsic laryngeal muscles. (From Hirano and Ohala, 1969)

the action potentials are obtained than do mono-polar electrodes.

Figure 5 gives a block diagram of the experimental apparatus used for gathering and processing the data. A detailed description of the characteristics of the EMG pre-amplifier and calibrator is given by Hubler (1967). The computer program used in averaging the data is described by Fromkin and Ladefoged (1966) and by Harshman and Ladefoged (1967). The F_o extraction system was built at the Speech Transmission Laboratory, Stockholm, in accordance with a design of Arne Risberg. It was modified by the addition of an interval-to-voltage converter (McKinney 1965). The frequency response of the data-gathering portion of the system was limited by the tape recorder which was flat (to within -3 dB) from 70 to 10 kHz. In the data processing phase the response was limited by the oscillomink and was flat (to within -3 dB) from DC to 1 kHz, or by the computer processing, which except for the initial high-pass filtering of the signal at 100 Hz (to remove possible movement artefacts) processes several signals (by adding them and subsequently "smoothing" them) in a way that is not easily quantifiable with respect to frequency response.

The EMG activity was evaluated in conjunction with the audio signal and the pitch curves on the basis of either oscillomink tracings of the "raw" electromyographic signals or computer-averaged EMG signals (or both). Computer averaging was useful when the activity of a given muscle during contraction was visually difficult to separate from its background activity, as was usually the case with the lateral cricoarytenoid, but hardly ever true in the case of the cricothyroid.

Meaningful comparison of the amplitudes of EMG records is restricted to those that have been obtained from a single insertion site during one recording session. It is not possible to say on the basis of EMG records alone whether muscle A contributed more to such-and-such a gesture than muscle B. Nor can we say meaningfully that muscle A in subject X showed more activity than the same muscle in subject Y. This should be borne in mind in considering the results presented below.

Results and Discussion

Figures 6 through 16 present some of the "raw" and processed EMG data obtained in this study.

First, it is quite obvious that the laryngeal muscles particpate actively in F_o control. A careful comparison of the records, for example those of the cricothyroid in Figures 8a and 8b which have the same phonetic "segments" but which differ only in that the word "Bob" in 8b has a terminal rise in pitch whereas that in 8a does not, proves that the differences in muscle action potentials are related to the pitch variations, not to segmental gestures. The cricothyroid is

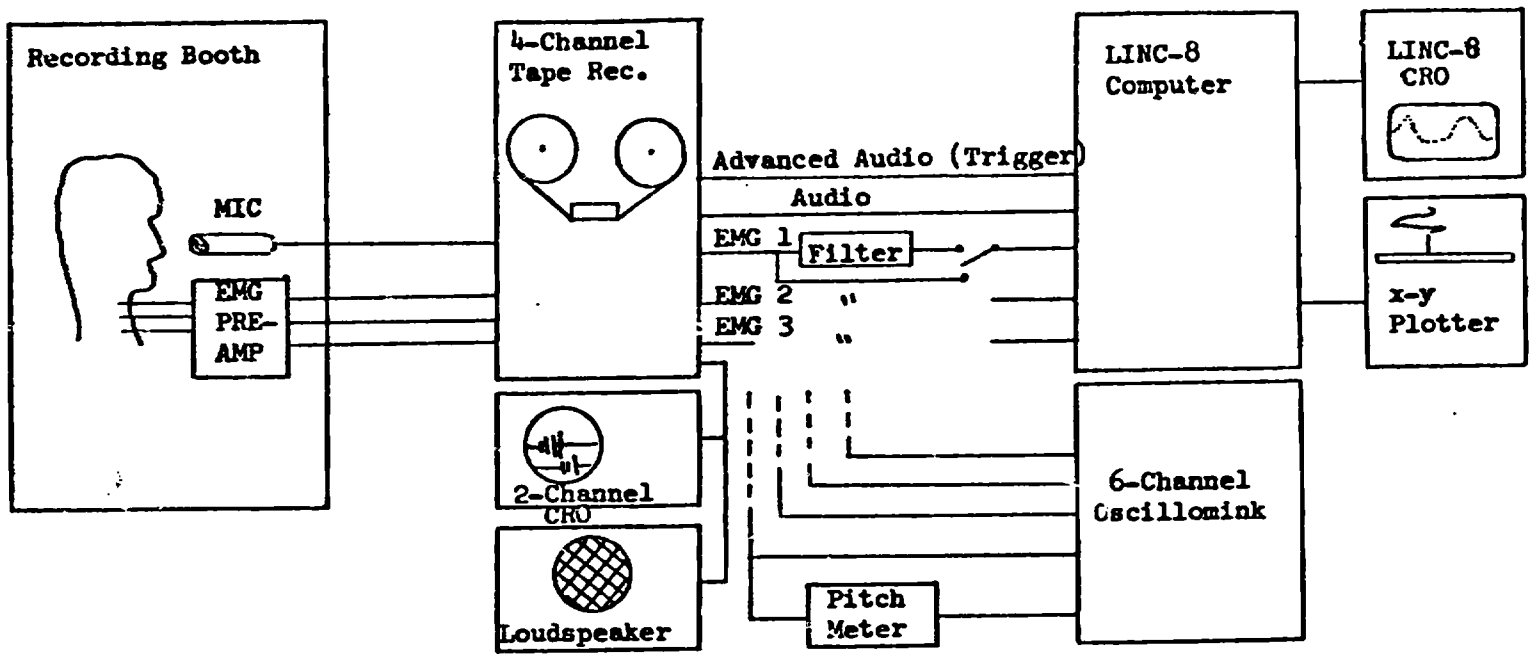


Figure 5. Schematic diagram of experimental apparatus.

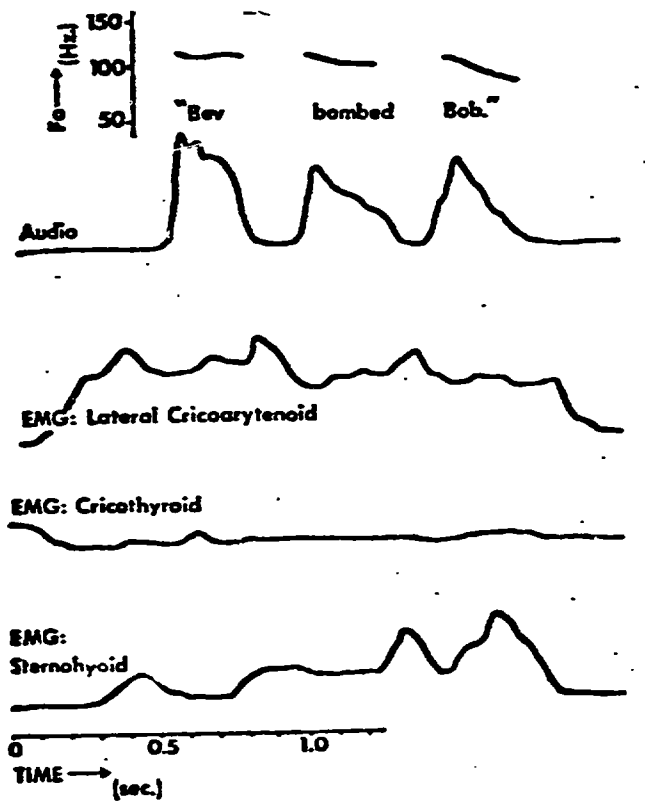


Figure 6a. Subject: JO; Signals represent 25 tokens which have been rectified, averaged and smoothed.

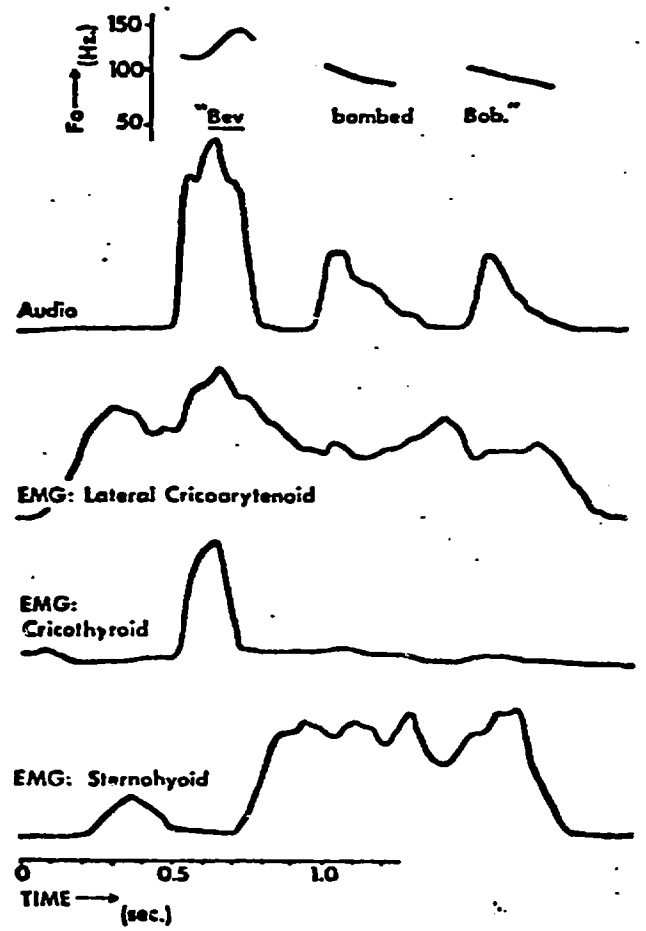


Figure 6b. As in Figure 6a, but with accent on "Bev".

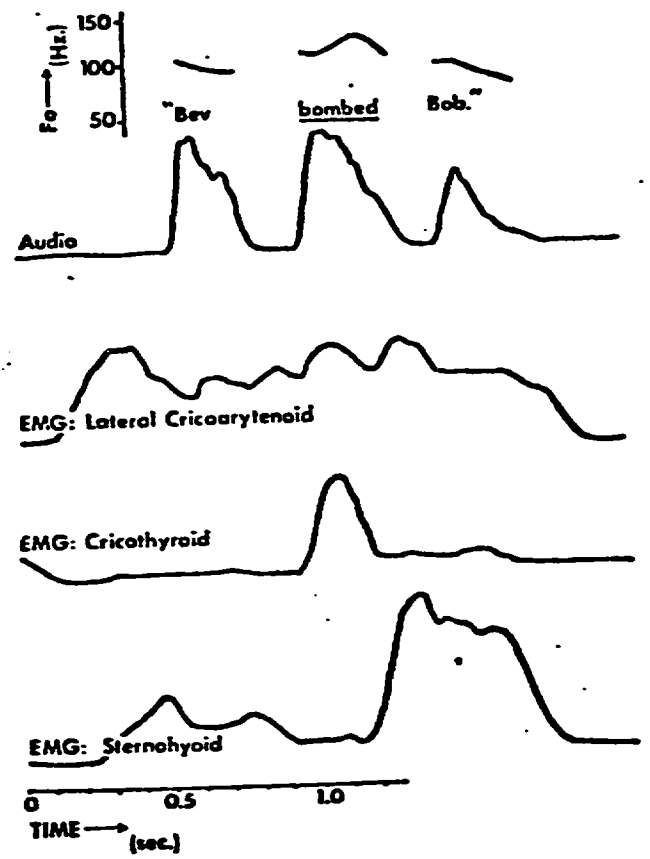


Figure 6c. As in Figure 6a, but with accent on "bombed".

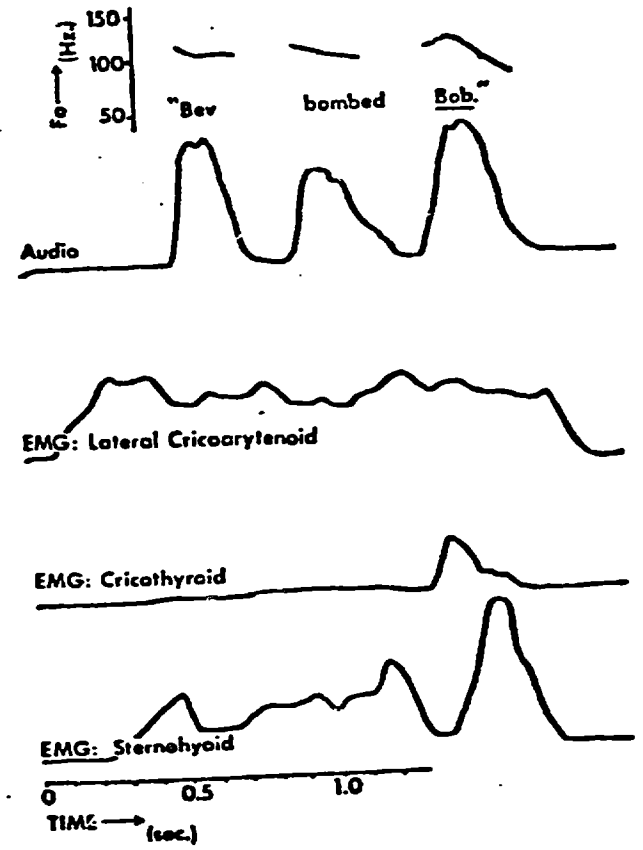


Figure 6d. As in Figure 6a, but with accent on "Bob".

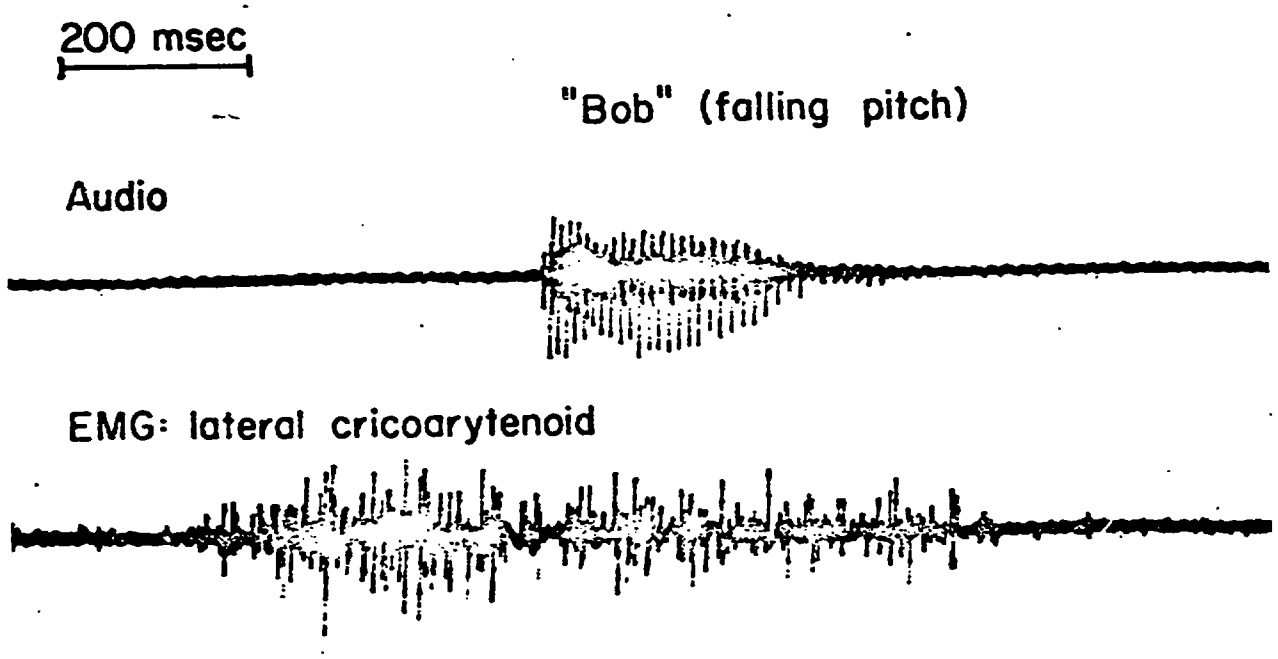


Figure 7a. Subject: JO; unprocessed signals.

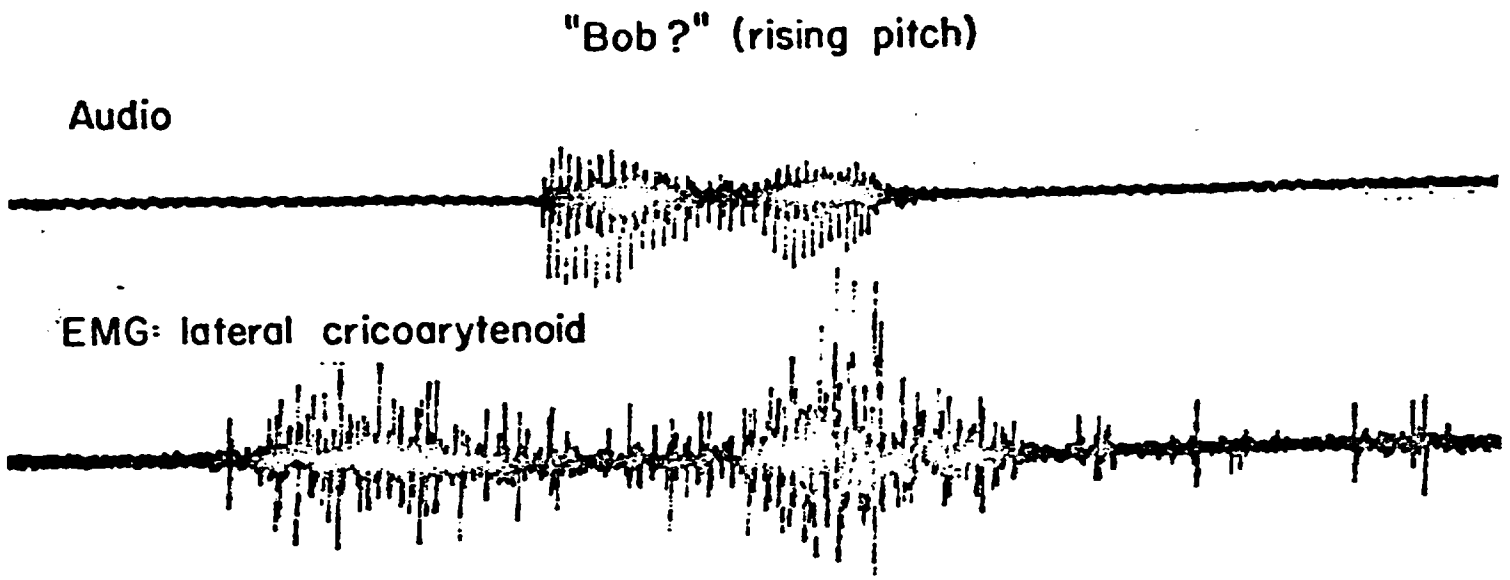


Figure 7b. As in Figure 7a.

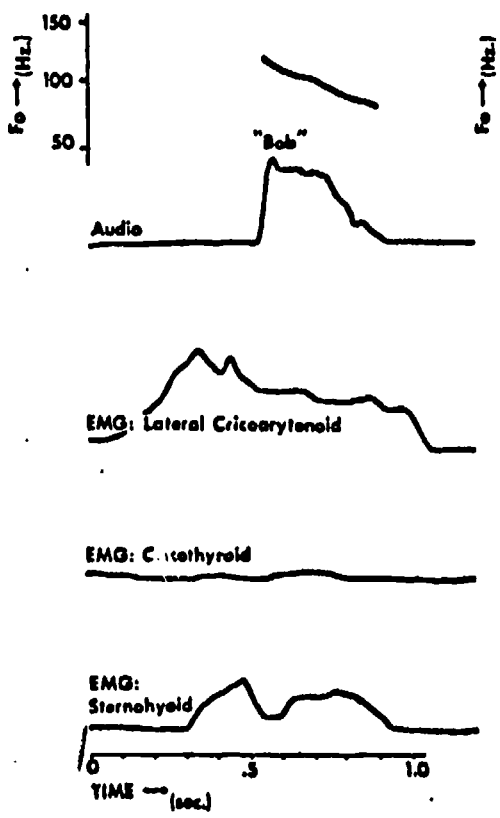


Figure 8a.

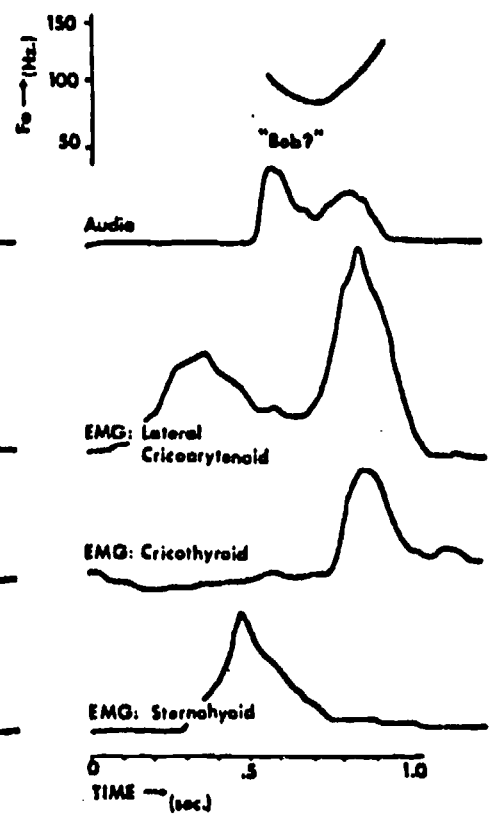


Figure 8b.

Subject: JO; parameters as in Figure 6a.

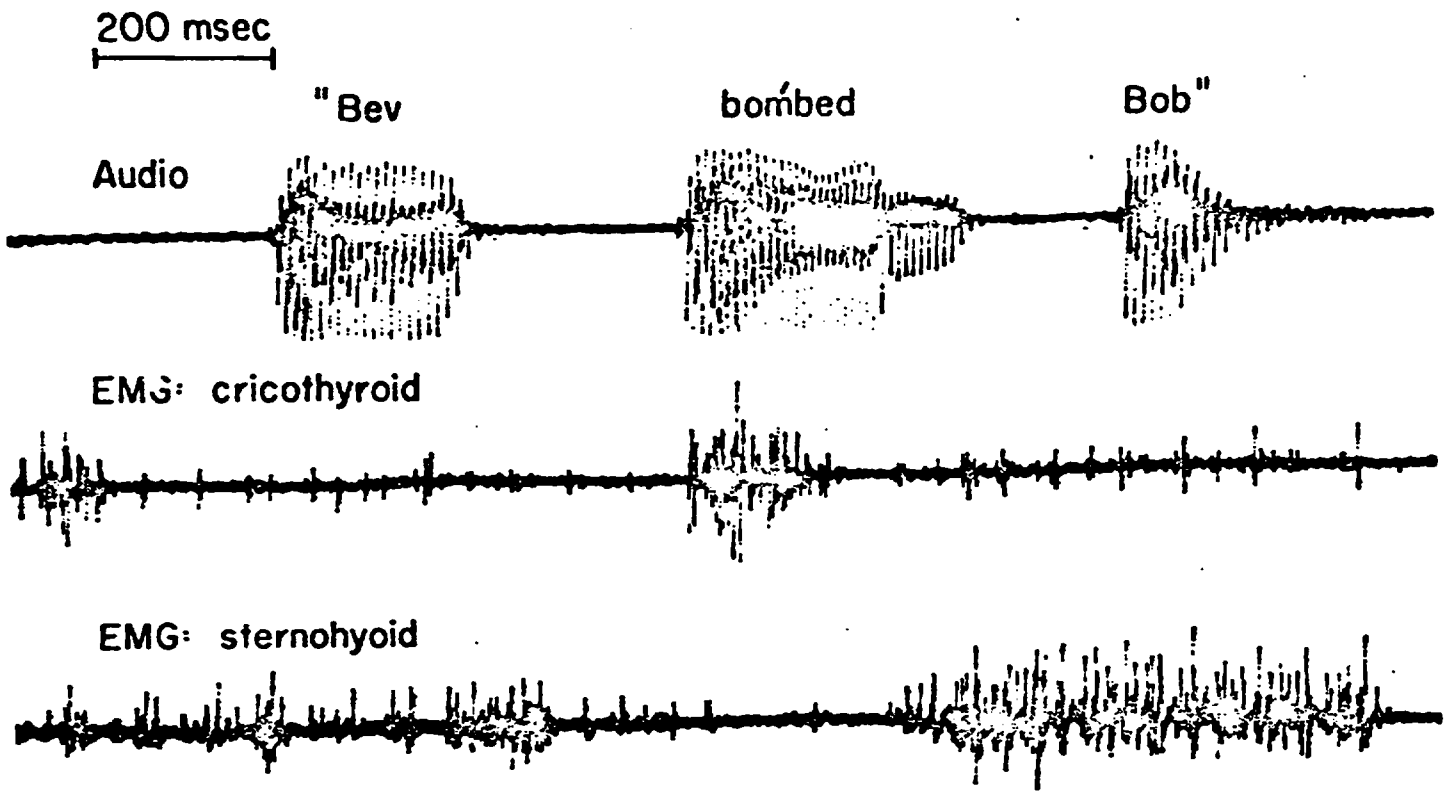


Figure 9. Subject: JO; unprocessed signals.

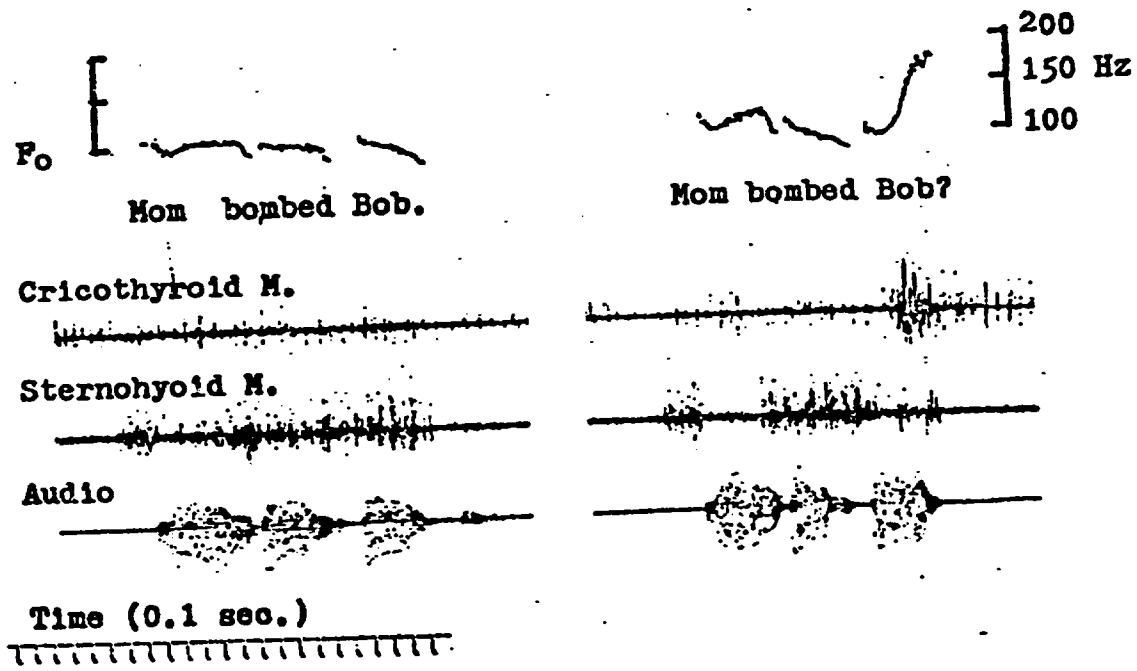


Figure 10a. Figure 10b.

Subject: JO; unprocessed signals.

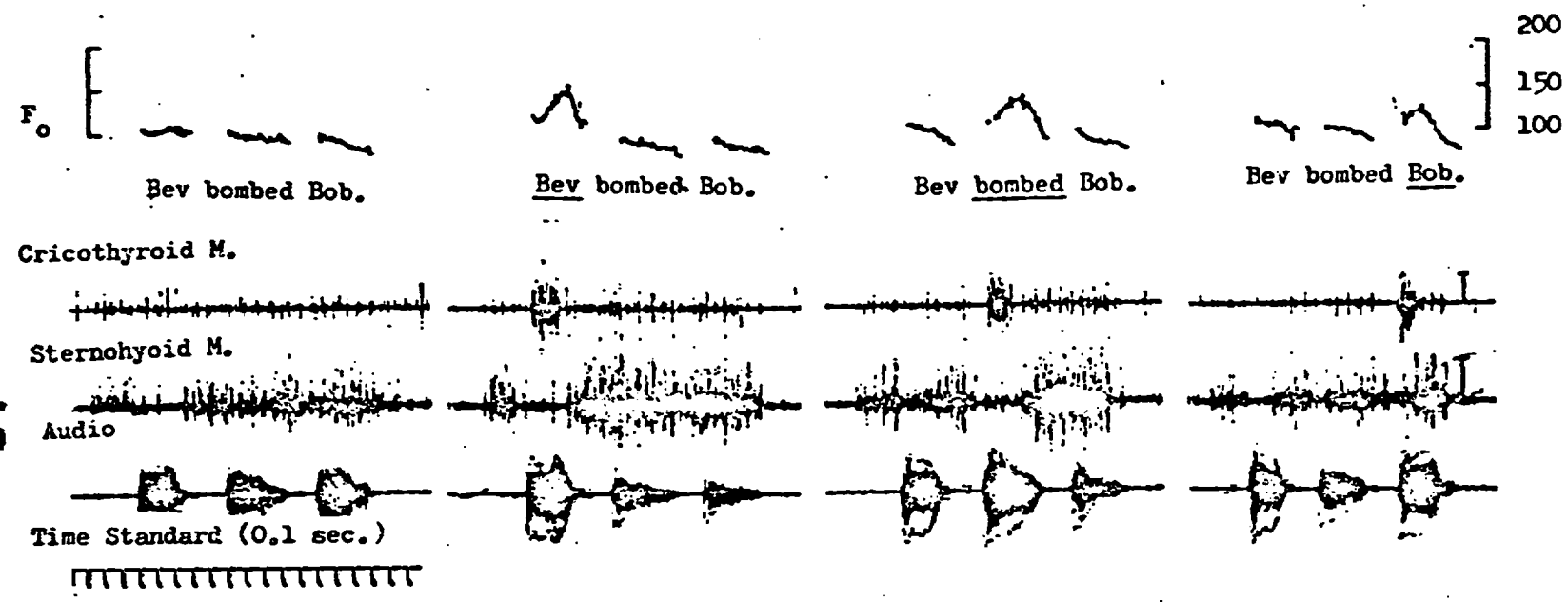


Figure 11a. Figure 11b. Figure 11c. Figure 11d.

Subject: JO; unprocessed signals.

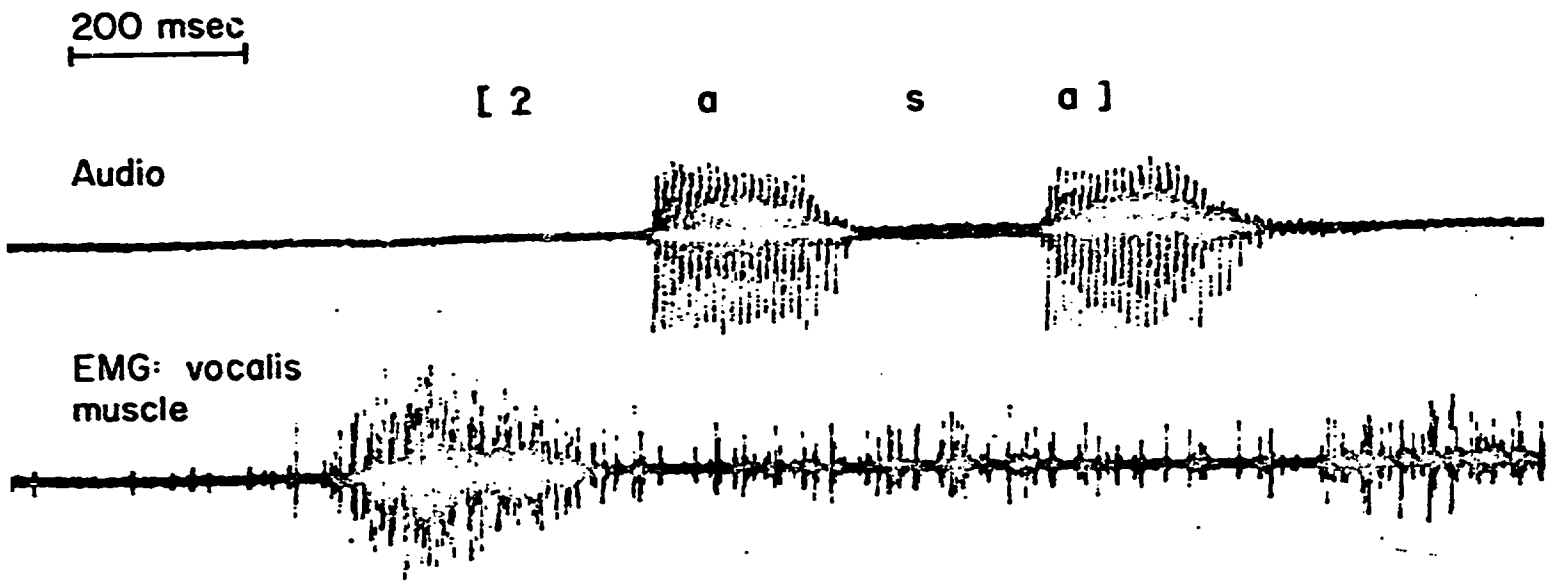


Figure 12. Subject: JO; unprocessed signals.

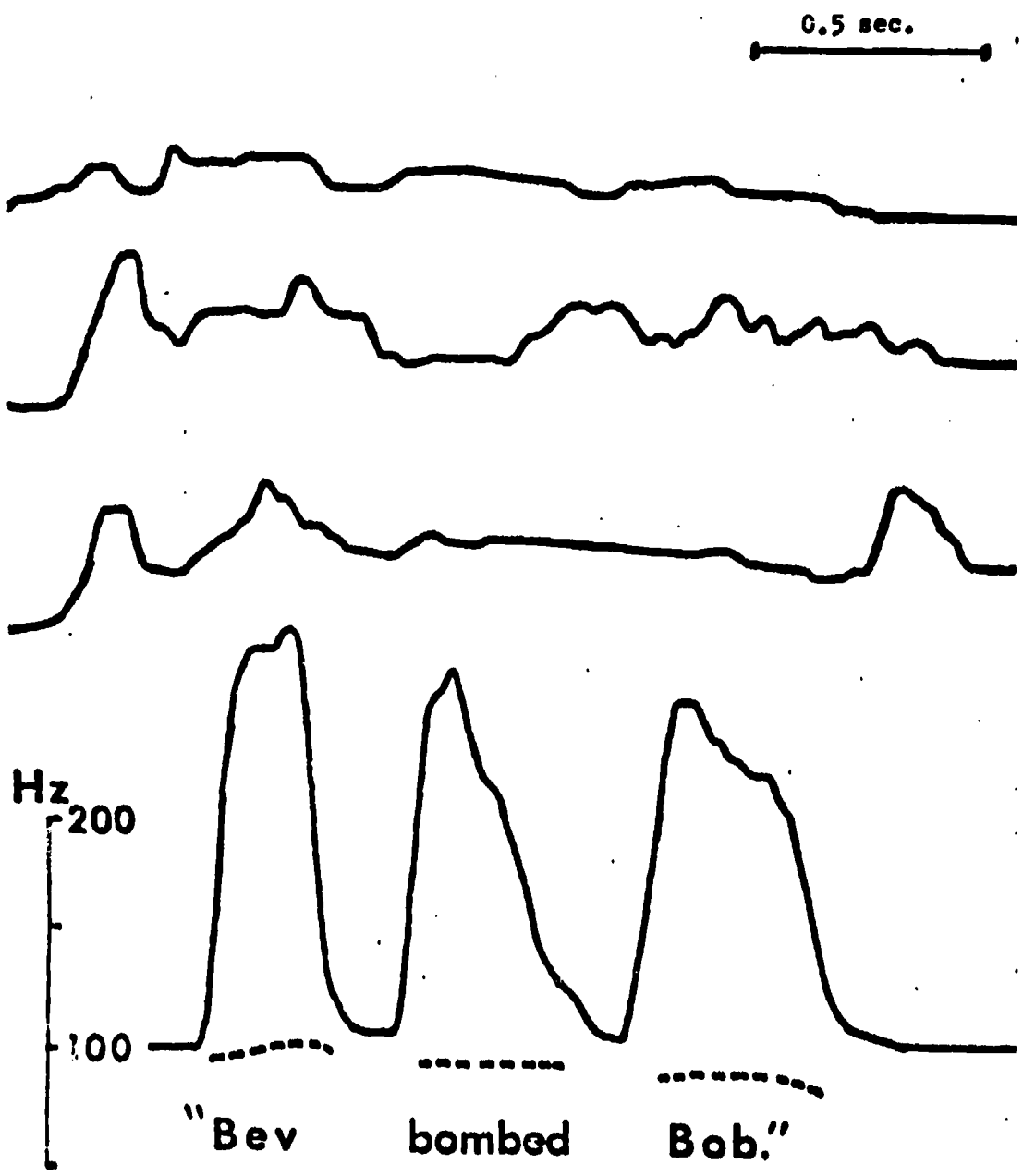


Figure 13. Subject: WV; processed signals as in Figure 6a; from top to bottom: cricothyroid muscle, lateral cricoarytenoid muscle, vocalis muscle, audio, and fundamental frequency (dotted line).

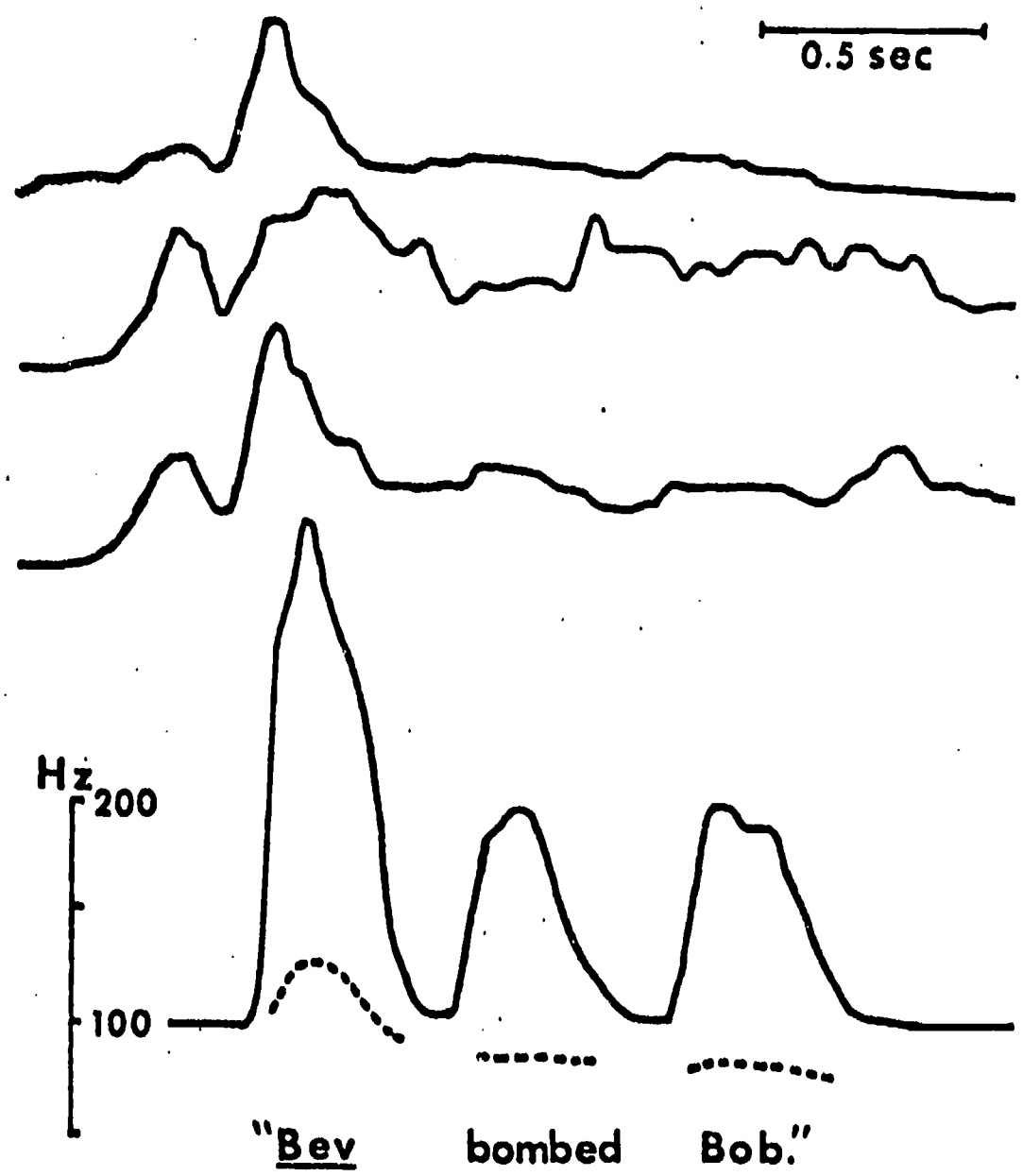


Figure 14. As in Figure 13.

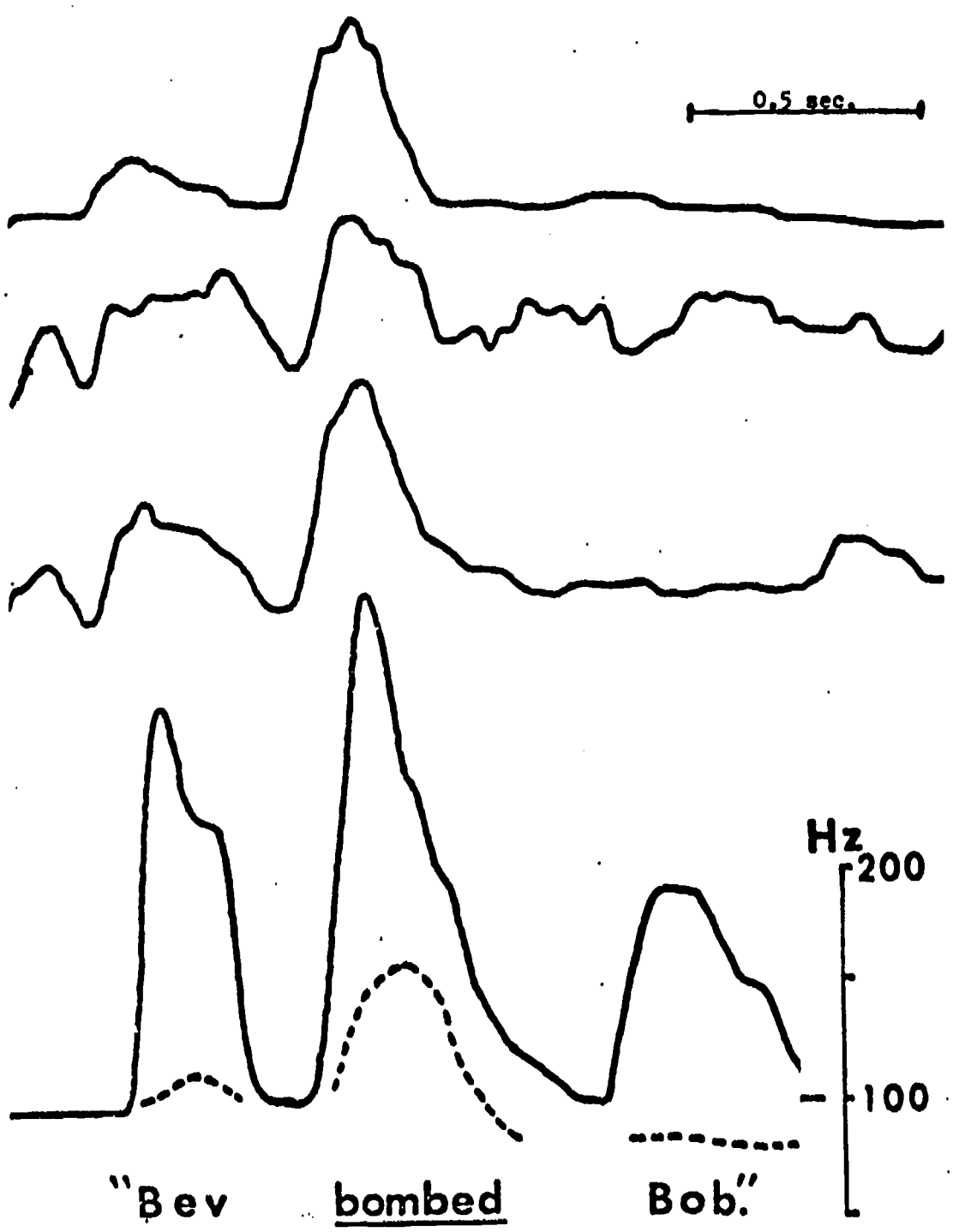


Figure 15. As in Figure 13.

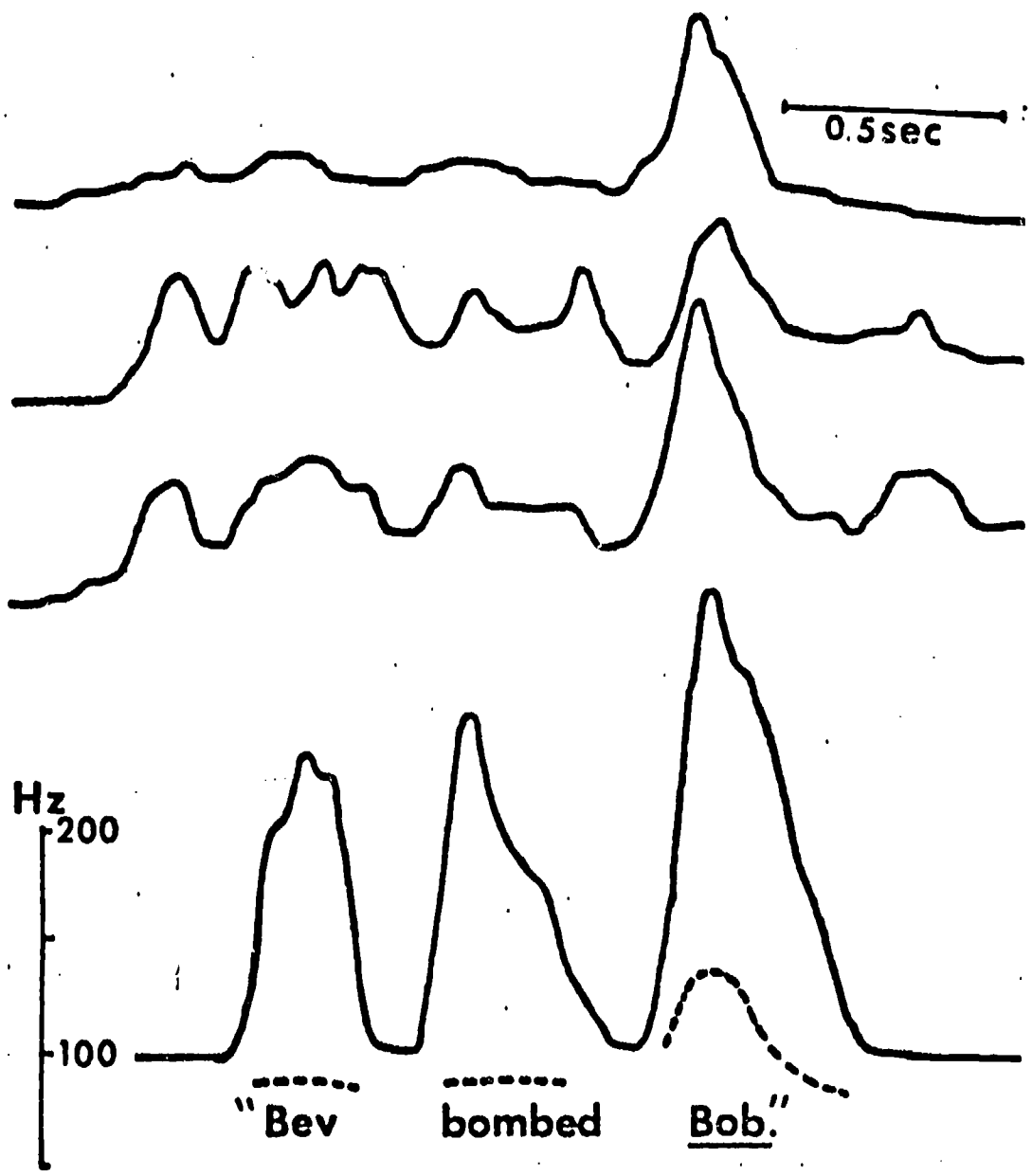


Figure 16. As in Figure 13.

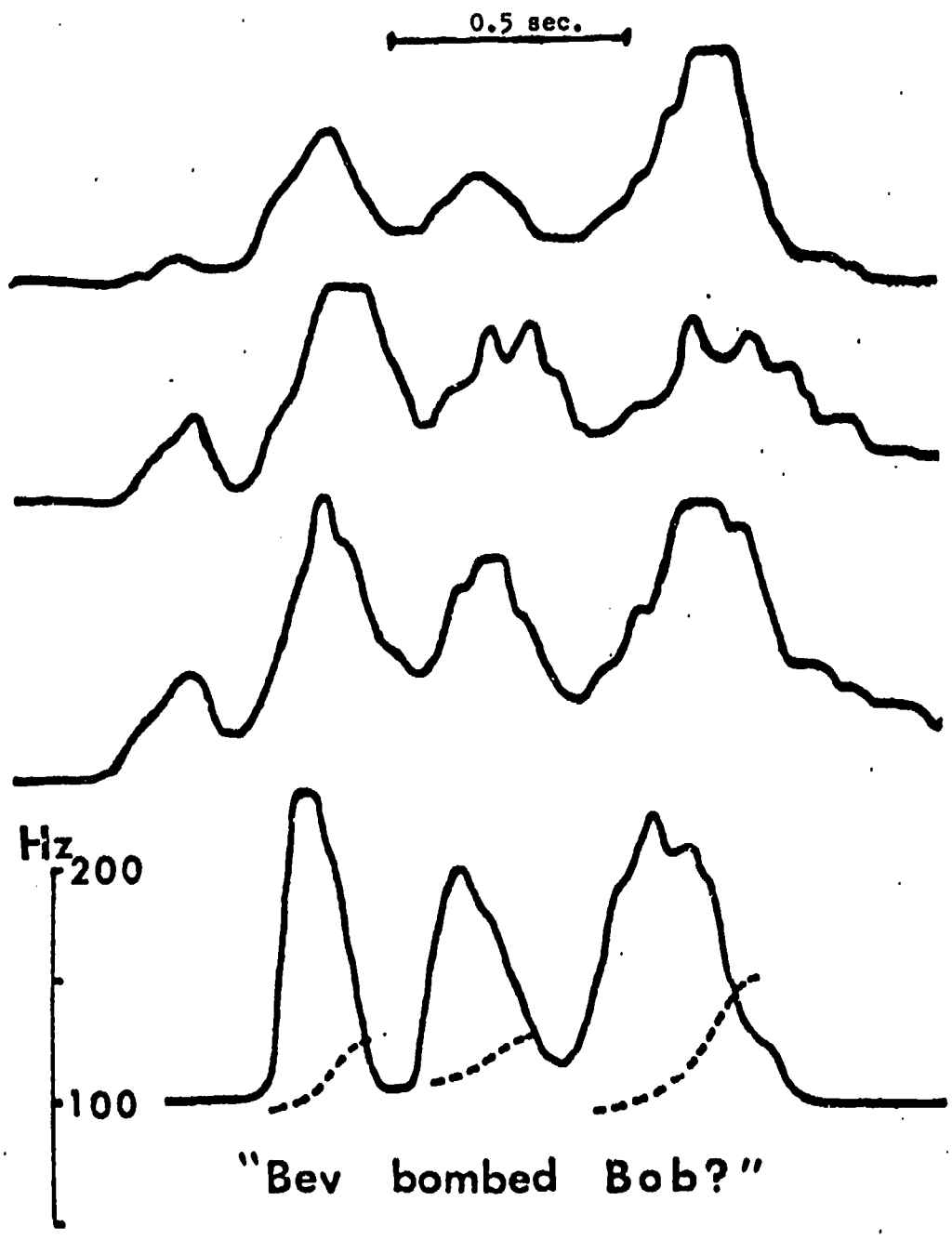


Figure 17. As in Figure 13.

relatively inactive, in fact, except when there is a rise in pitch. On the other hand, the lateral cricoarytenoid is active immediately before the onset of and throughout periods of voicing, but can be seen to increase in activity during a rise in pitch, as, e.g., in Figures 7a and 7b. The sternohyoid shows intermittant activity, some of it clearly associated with non-prosodic gestures, and some with lowering pitch, e.g, in Figures 11a-d, the sternohyoid shows activity at the beginning of the sentences (perhaps for the jaw opening in "Bev") as well as during lowering of pitch.

Contrary to the findings of Katsuki (1950), Zenker and Zenker (1960), Perkins and Yanagihara (1968), and Wenrick (1931) there was no indication in this investigation that an increase in the activity of the cricothyroid muscle is associated with low pitch. Katsuki found this pattern only when he used surface electrodes and thus the signals obtained could easily reflect contamination from the sternohyoid, which, as mentioned, does show increased activation for low pitch. This could also explain the findings of Wenrick who used surface stimulation on the larynx, assuming that only the cricothyroid muscle would be affected by the electric pulse applied lateral to the cricothyroid space. Since Zenker and Zenker as well as Perkins and Yanagihara used needle electrodes in their studies, it is possible that they, too, were actually recording from the sternohyoid muscle in those cases where they found increased activity for low pitch. Hirano and Ohala (1969) have shown that the depth of insertion of a needle electrode is of critical importance in order to obtain recordings of the cricothyroid muscle and not the sternohyoid (see Figure 18). And it is quite easy for a stiff needle electrode to slip out of its original placement during the course of an electromyographic recording session. Further evidence that Zenker and Zenker were recording from the sternohyoid is the report by Zenker (1964) that they found increased activity in what they claimed to be the cricothyroid muscle for such gestures as jaw opening and back tilting of the head, two gestures more reasonably involving activation of the sternohyoid but not of the cricothyroid.

Second, the cricothyroid and lateral cricoarytenoid show increases in activity not only for terminal rises in pitch, where it is not disputed even by Lieberman that the laryngeal muscles cause the pitch rise, but also for pitch rises anywhere else in the breath-group, where he claims that the pitch rises result primarily from increases in subglottal pressure. Compare, for example, Figures 14, 15, and 16 with Figure 17. Further, these increases in muscle action potentials are of the same order of magnitude no matter where the pitch rise occurred within the breath-group. If it is admitted that the laryngeal muscles are the prime cause of pitch rise in one environment, it is hard to avoid the conclusion that when they show the same degree of activity elsewhere they are the prime cause of the pitch rise there too.

Third, given the common observation that the larynx moves up and down in conjunction with variations in pitch, it is not surprising to find that the sternohyoid, which is capable of lowering the entire larynx-hyoid apparatus, shows increased activity when pitch is lowered, as can be seen in Figures 6a through 6d, 10a and 10b, and 11a through 11d. But this again provides no support for Lieberman's contention that the terminal drop in F_o at the end of "unmarked" breath-groups is due to the falling subglottal pressure. And, as with the muscles active in raising pitch, the sternohyoid exhibits this pattern no matter whether the drop in pitch occurs at the end of a breath-group, as it does in Figure 10a or if it is well before the end, as it is in Figure 10b. The remaining activity of the sternohyoid is probably due to its participation in movements of the jaw opening and tongue retraction, although this has yet to be firmly established.

Figures 14 through 17 show significant increases in vocalis activity concomitant with increases in F_o, although it is active as well for gestures not associated with pitch, e.g., for glottal stops, as in Figure 17. These findings agree with those of Faaborg-Andersen (1957 and 1965). An investigation of singing in which either intensity or F_o were held constant while the other parameter changed revealed that in chest register the vocalis participates both in increases in pitch and intensity -- see Hirano, Ohala, and Vennard (1969). The large increase in the vocalis muscle's activity during pitch rises on emphasized words suggests that the glottal resistance increases at these points. This being the case it is possible that some part of the increased subglottal pressure encountered on emphasized words is due to this increased glottal resistance and not entirely due to increased respiratory effort. This is not surprising and is in essential agreement with the above-mentioned studies on animal larynges (Koyama, et al. 1969) as well as the aerodynamic studies of speech activity (Ladefoged 1963, Broad 1968). But this provides no support for the claim underlying Lieberman's assertions about the dependency of F_o on subglottal pressure. His assertions rely on the assumption that subglottal pressure is itself not dependent on laryngeal adjustment. To the extent that subglottal pressure is itself partly determined by laryngeal adjustment it is logically not possible to say that F_o in speech is determined by subglottal pressure and not all or only negligibly by laryngeal adjustment.

Only the interarytenoid and the posterior cricoarytenoid muscles showed no activity correlating with pitch variations. (In fact, the interarytenoid activity did occur with pitch rises, at the extreme high end of the subject's pitch range -- well beyond that used for speech. Cf. Ohala, Hirano, and Vennard, 1968). Rather these muscles were more active with segmental or respiratory gestures.

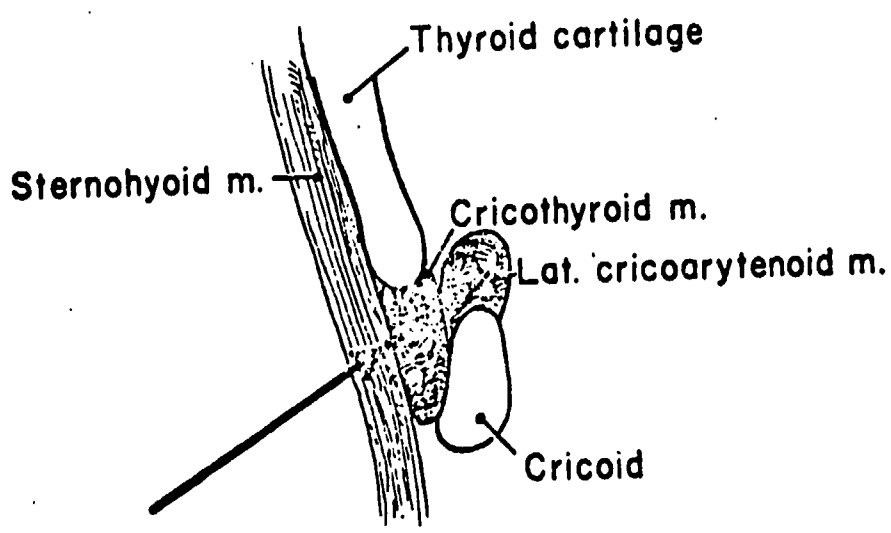


Figure 18. Schematic diagram showing the possibility of having an electrode aimed at the cricothyroid muscles slip back into the sternohyoid muscle instead. (From Hirano and Ohala, 1969)

STUDY TWO: CALIBRATION OF EFFECT OF SUBGLOTTAL PRESSURE ON PITCH

A calibration of the effects of either laryngeal adjustment or subglottal pressure on F_o of phonation requires holding one of these constant and varying the other. It is easier to hold the laryngeal adjustment constant and vary the subglottal air pressure, rather than the other way around, since laryngeal adjustment is difficult to quantify anyway. The procedure repeated that used by Isshiki (1959), Ladefoged (1963), and Öhman and Lindqvist (1966a). The experiment involves having a subject maintain a steady pitch and then pushing on his chest or abdomen at unexpected moments. The pushes cause rapid involuntary increases in pitch. The assumption underlying this experiment is that there will be insufficient time for the subject to reflexly adjust his laryngeal muscles and that the laryngeal adjustment can be taken as remaining constant. One can then measure the increase in subglottal pressure and the concomitant increase in F_o and calculate the effect of increases in subglottal pressure on the pitch in Hz/cm. aq. In the present study, as in that of Isshiki,* the activity of the laryngeal muscles (in the present study, the cricothyroid and lateral cricoarytenoid) was sampled in order to make sure that their tension did not change during the pushes.

Subject JO from the previous study was the only subject in this study. Subglottal pressure was obtained from a modified spinal needle inserted in the larynx between the thyroid and cricoid cartilage, which was connected via a short plastic tube to a Greer photoelectric air pressure transducer. The frequency response of the system was flat within -3 dB from DC to 25 Hz. The subject was seated comfortably in a chair with his eyes closed and attempted to maintain a given pitch while phonating the open vowel /a/. At unexpected moments he received a slight push on the chest or abdomen. Various intensities, various pitches and various modes of phonation were used in calculating $\Delta F_o/\Delta P_s$. The pressure before a given push was subtracted from the peak pressure to give ΔP_s and the difference between the values for F_o at these two points provided ΔF_o. The initial value of F_o was plotted against the initial value of P_s and the peak F_o against the peak P_s, the slope of the line joining these two points then equalling $\Delta F_o/\Delta P_s$. The determination of any given P_s value (from the oscillomink write-out of the pressure) is estimated to be accurate to within ± 0.5 cm. aq. and the accuracy of the F_o measurements (from narrow band spectrograms) are estimated to be accurate to within 2 °/o of the measured value. These errors are compounded in determining the slopes of the lines and are likely to be non-negligible when the line segment is short since in that case the magnitude of error approaches that of the line itself.

Results and Discussion

Monitoring the activity of the two laryngeal muscles was found to be important because changes in their level of activity did sometimes

---

* Isshiki monitored the activity of the anterior cricothyroid by concentric unipolar needle electrodes inserted through the skin and the activity of the vocalis muscle by the same type of electrode inserted via the vocal tract (Isshiki, personal communication).

occur during a push, especially when the person doing the pushing stopped being random in the interval between pushes. Apparently the subject was able to follow the slight rhythm that was developing in the time of the pushes. When this happened there was a decrease in the activity of the cricothyroid muscle which could have the effect of lowering the pitch at the moment of the push. Only the pressure peaks and F_o deviations that did not show any accompanying change in the activity of these two laryngeal muscles were used in the calculations for finding the influence of subglottal pressure variations on the F_o.

In figures 19 through 32 are plotted the changes in frequency against the changes in pressure (as described above) for 161 pushes on the subject's chest and abdomen for different conditions of phonation. As is evident, the only large difference in the slopes of the lines is that between falsetto and everything else, the slope being about 7-10 Hz/cm. aq. for falsetto and about 2-4 Hz/cm. aq. otherwise, with the value for the pitch range used in speech being only 2-3 Hz/cm. aq. None of the other variations in the conditions of phonation seem to make a systematic difference although this is very possibly due to the subject's lack of voice training and consequent inability to execute the required phonation conditions accurately, e.g., low intensity tended to be combined with breathy voice and high intensity with tense voice in a non-systematic way. Unfortunately at the time of running this study the relevance of the claims of Flanagan and Landgraf (1967 and 1969) and Flanagan (1968) had not been realized and consequently variations in vocal tract shape were not among the various conditions included.* These data, then, offer no evidence on the possibility of whether or not acoustic coupling between the vocal tract and the vocal cords could allow a given change in subglottal pressure to affect the F_o more when it occurred with a vocal tract whose first resonance (formant) approached that of the fundamental. Possibly the high value for $\Delta F_o / \Delta P_s$ during falsetto is due to this effect since the first resonance of the vocal tract for the vowel /a/ is near the fundamental in this case. However, from van den Berg and Tan's high value for $\Delta F_o / \Delta P_s$ for falsetto obtained from an excised larynx (see Table 1), one would expect such a high value to be due to the peculiar mode of phonation used in falsetto and not due merely to effects of acoustic coupling. An elucidation of this point awaits further experimentation.

These values compare favorably with those of Isshiki, Ladefoged, and Öhman and Lindqvist, all of whom worked with living subjects, and are at least in the range of values reported by van den Berg and Tan, Furukawa, and Anthony, who worked with excised larynges, the differences being attributable to individual variation or to the different experimental conditions (living subject versus excised larynx). However, the difference between the values found in this study and the values derived by Lieberman (1967a) from running speech, namely 16-22 Hz/cm. aq. are too large to be attributed to individual variation. This is not surprising, however, since Lieberman did not do any experiment of the

---

* However, see Ohala and Ladefoged (1969) and Hixon, Mead, and Klatt (1970). Both of these studies failed to find any support for the claims of Lieberman or the theoretical predictions of Flanagan and Landgraf regarding the influence of P_s on fundamental frequency.

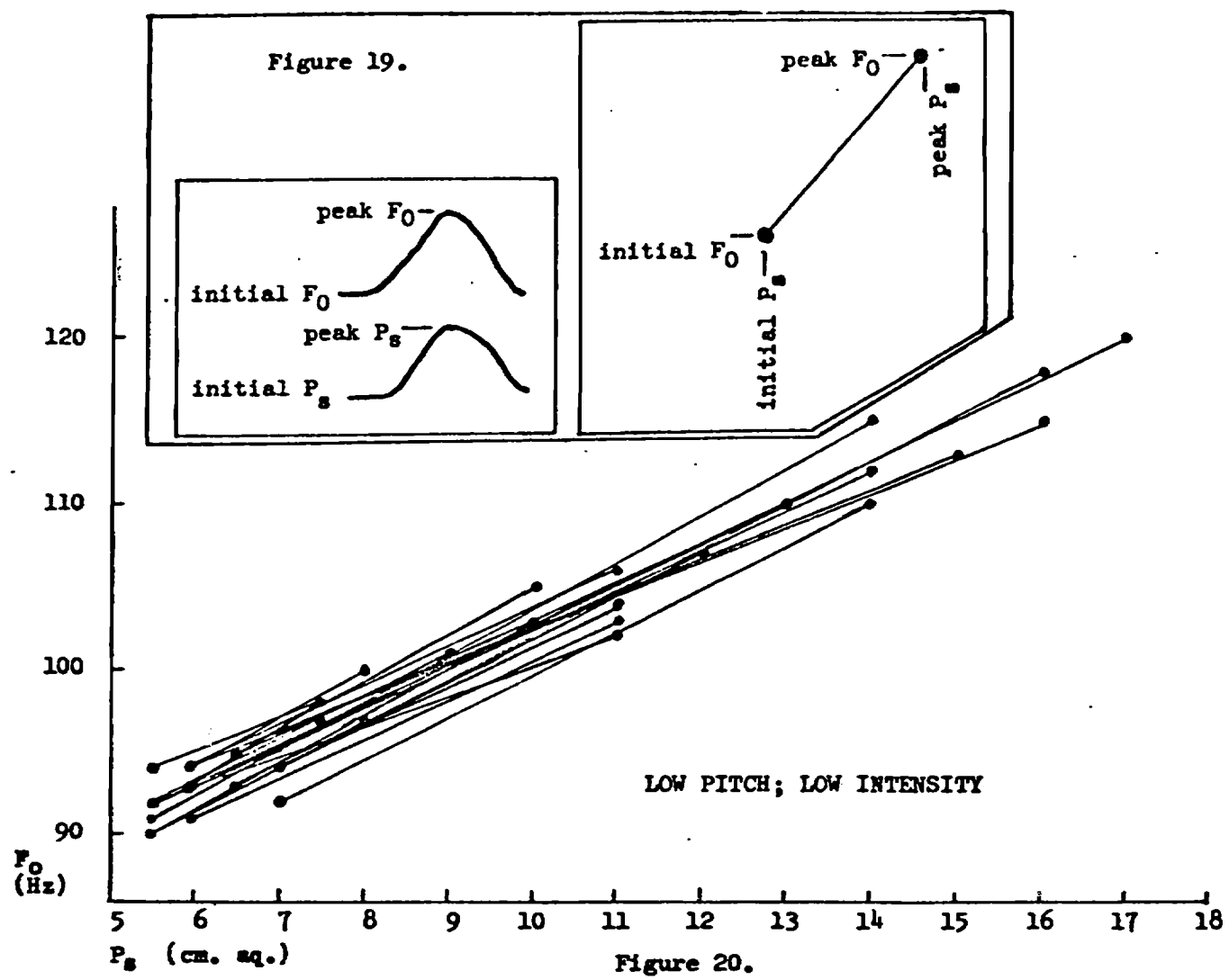


Figure 19.

Figure 20.

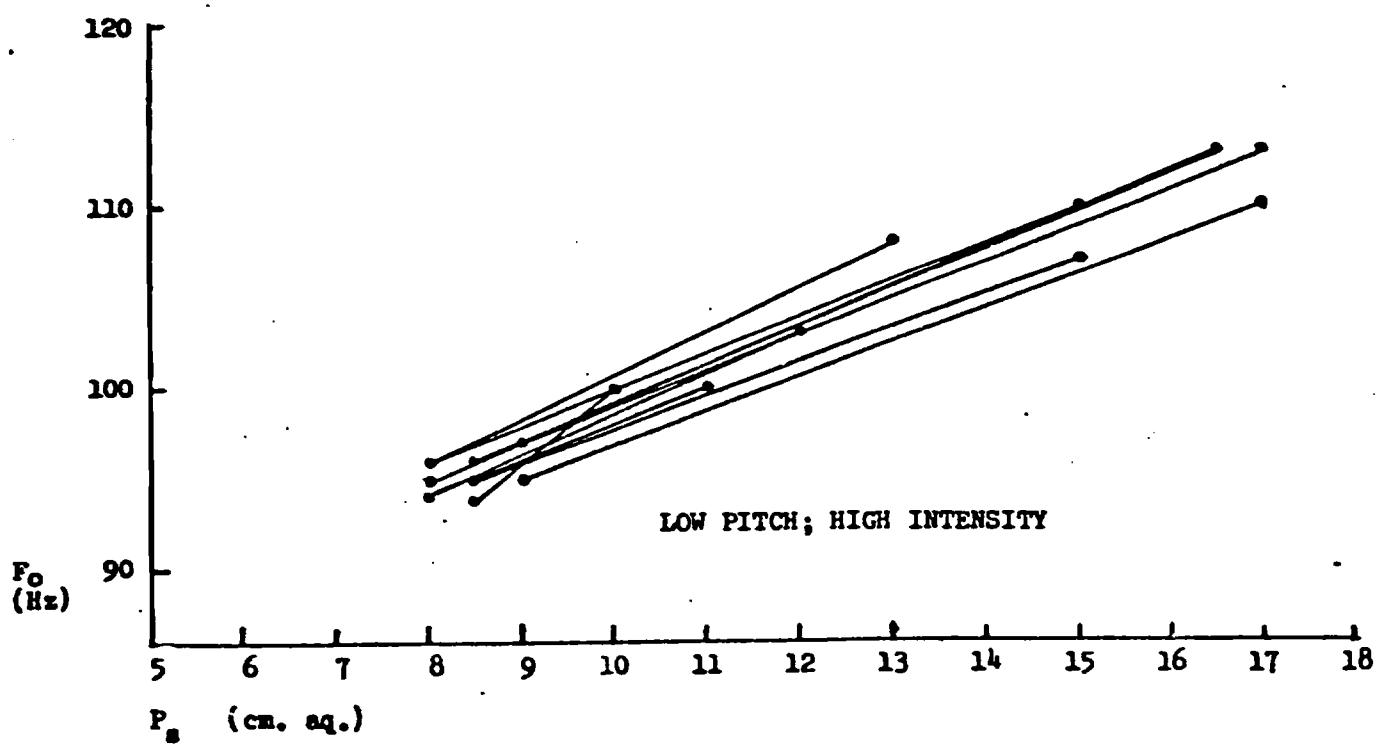

120
110
100
90
F_o (Hz)
LOW PITCH; HIGH INTENSITY
5
6
7
8
9
10
11
12
13
14
15
16
17
18
P_s (cm. aq.)

Figure 21.

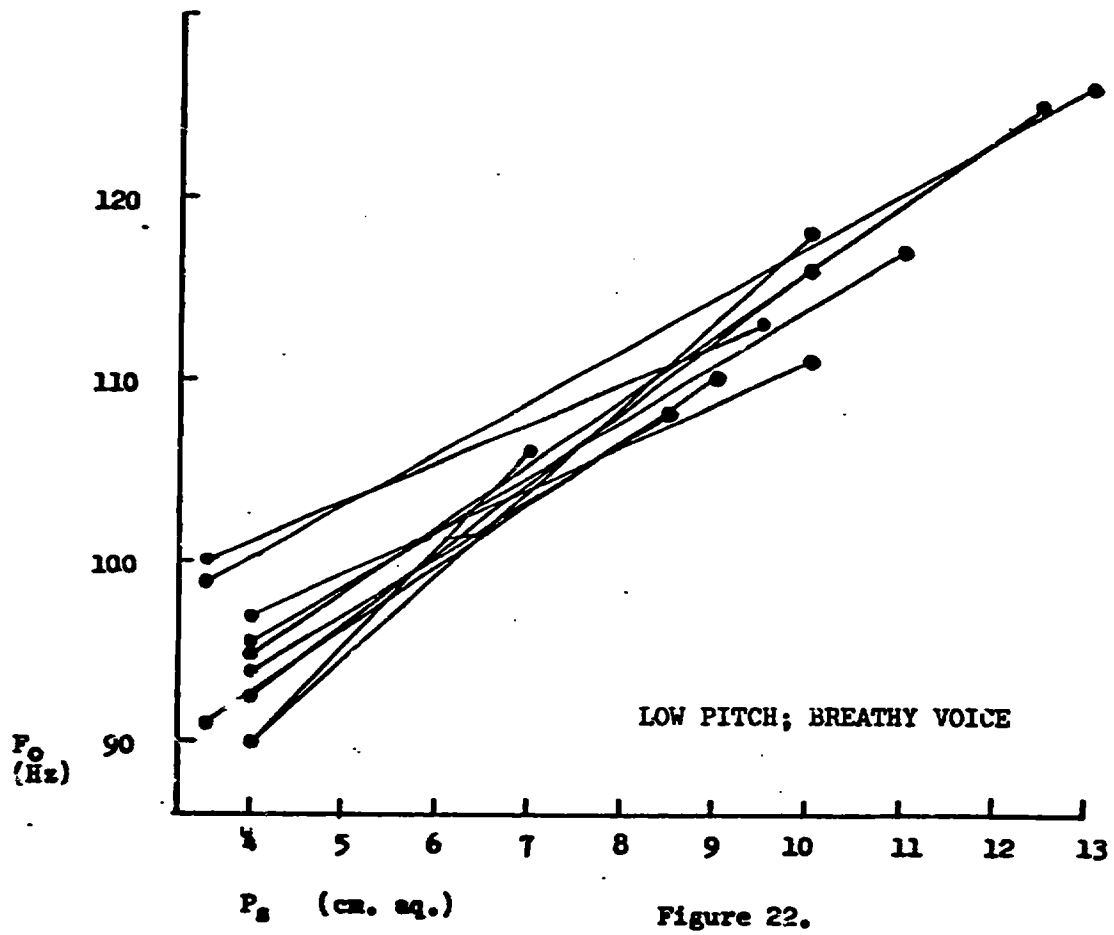

120
110
100
90
F_o (Hz)
LOW PITCH; BREATHY VOICE
4
5
6
7
8
9
10
11
12
13
P_s (cm. aq.)

Figure 22.

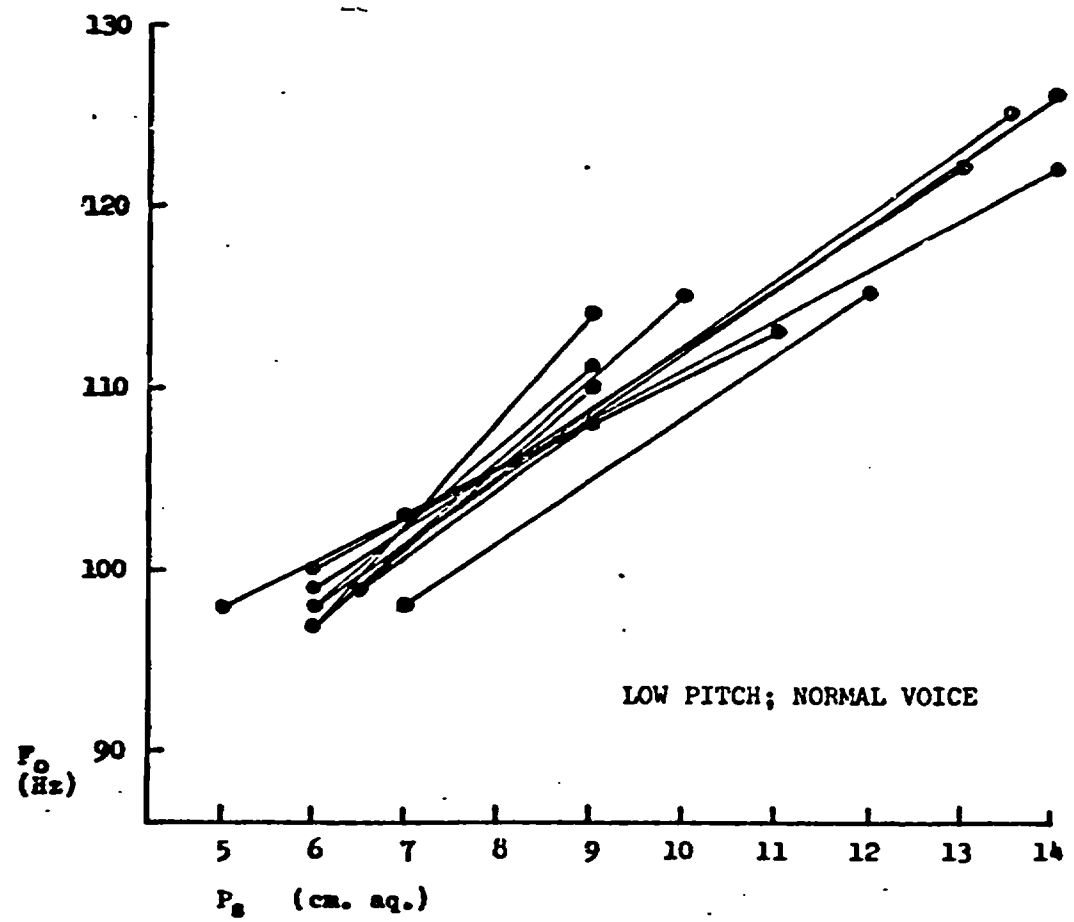

130
120
110
100
90
F_o (Hz)
LOW PITCH; NORMAL VOICE
5
6
7
8
9
10
11
12
13
14
P_s (cm. aq.)

Figure 23.

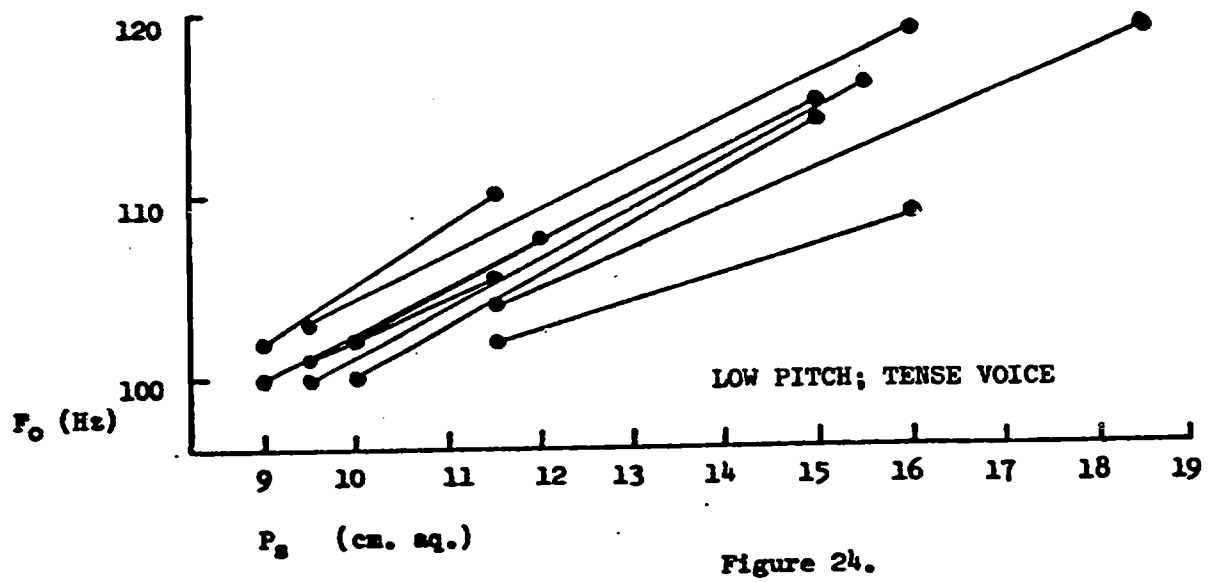

120
110
100
F_o (Hz)
LOW PITCH; TENSE VOICE
9
10
11
12
13
14
15
16
17
18
19
P_s (cm. sq.)

Figure 24.

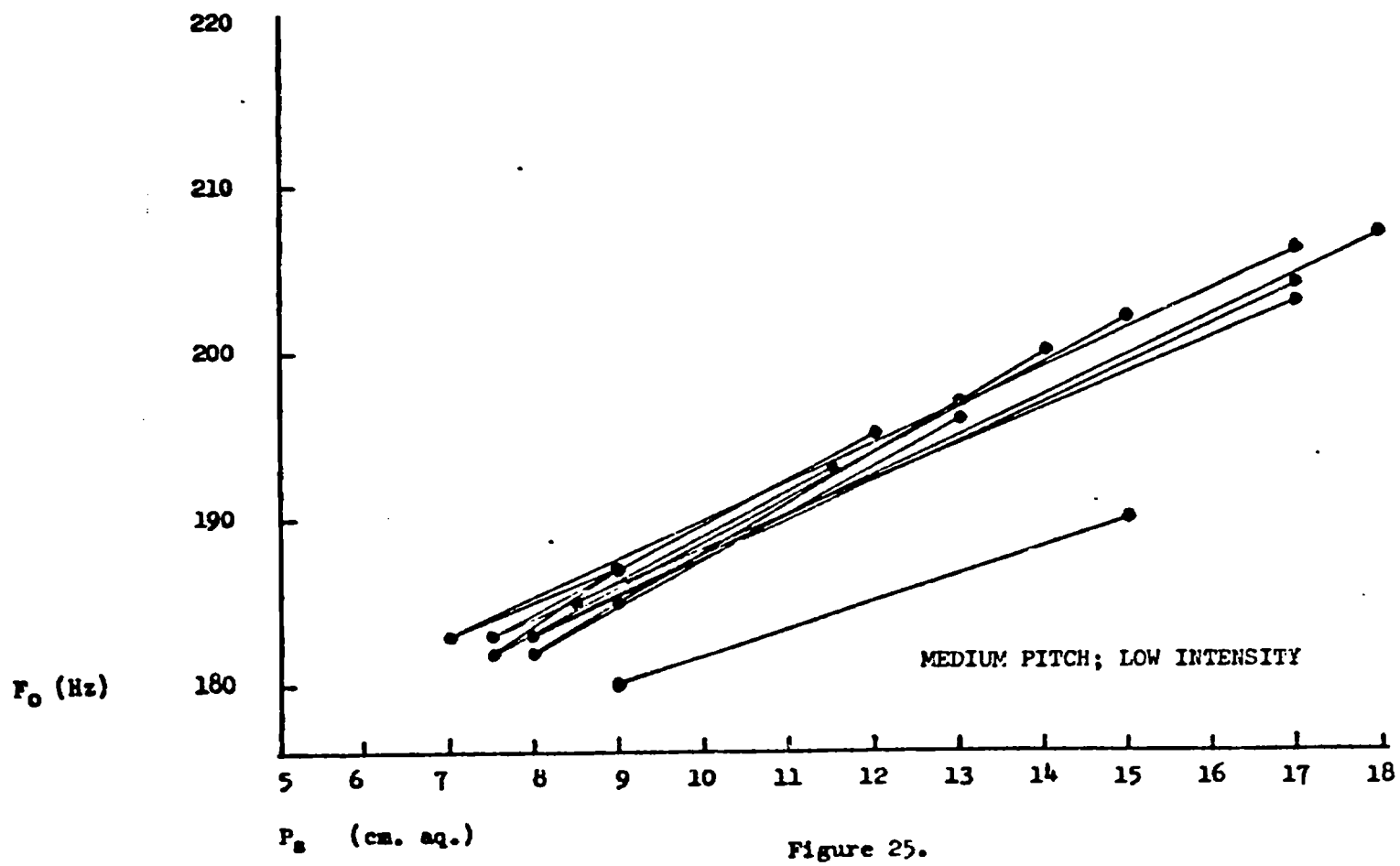

220
210
200
190
180
F_o (Hz)
MEDIUM PITCH; LOW INTENSITY
5 6 7 8 9 10 11 12 13 14 15 16 17 18
P_s (cm. aq.)

Figure 25.

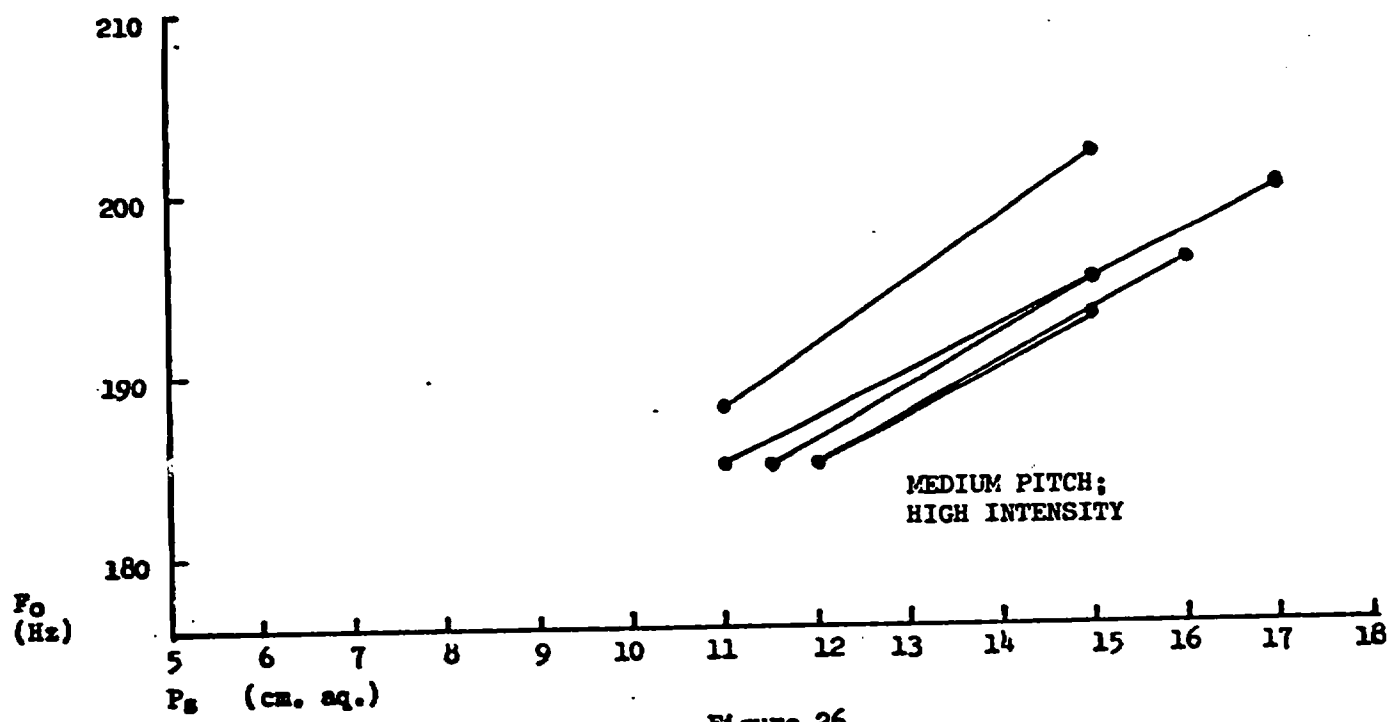

210
200
190
180
Fo
(Hz)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Ps (cm. aq.)
MEDIUM PITCH;
HIGH INTENSITY

Figure 26.

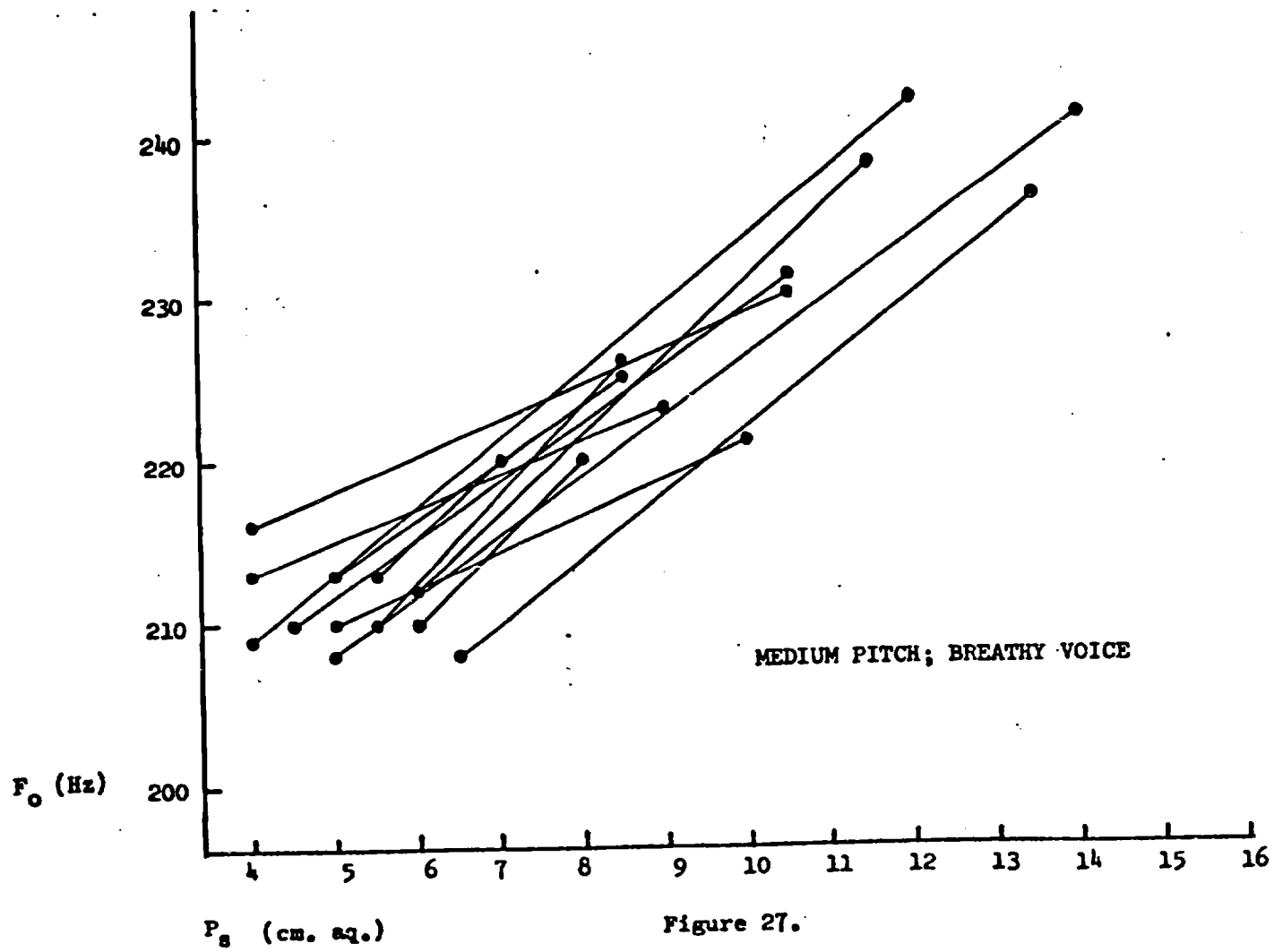

240
230
220
210
200
F_o (Hz)
4
5
6
7
8
9
10
11
12
13
14
15
16
P_s (cm. aq.)
MEDIUM PITCH; BREATHY VOICE

Figure 27.

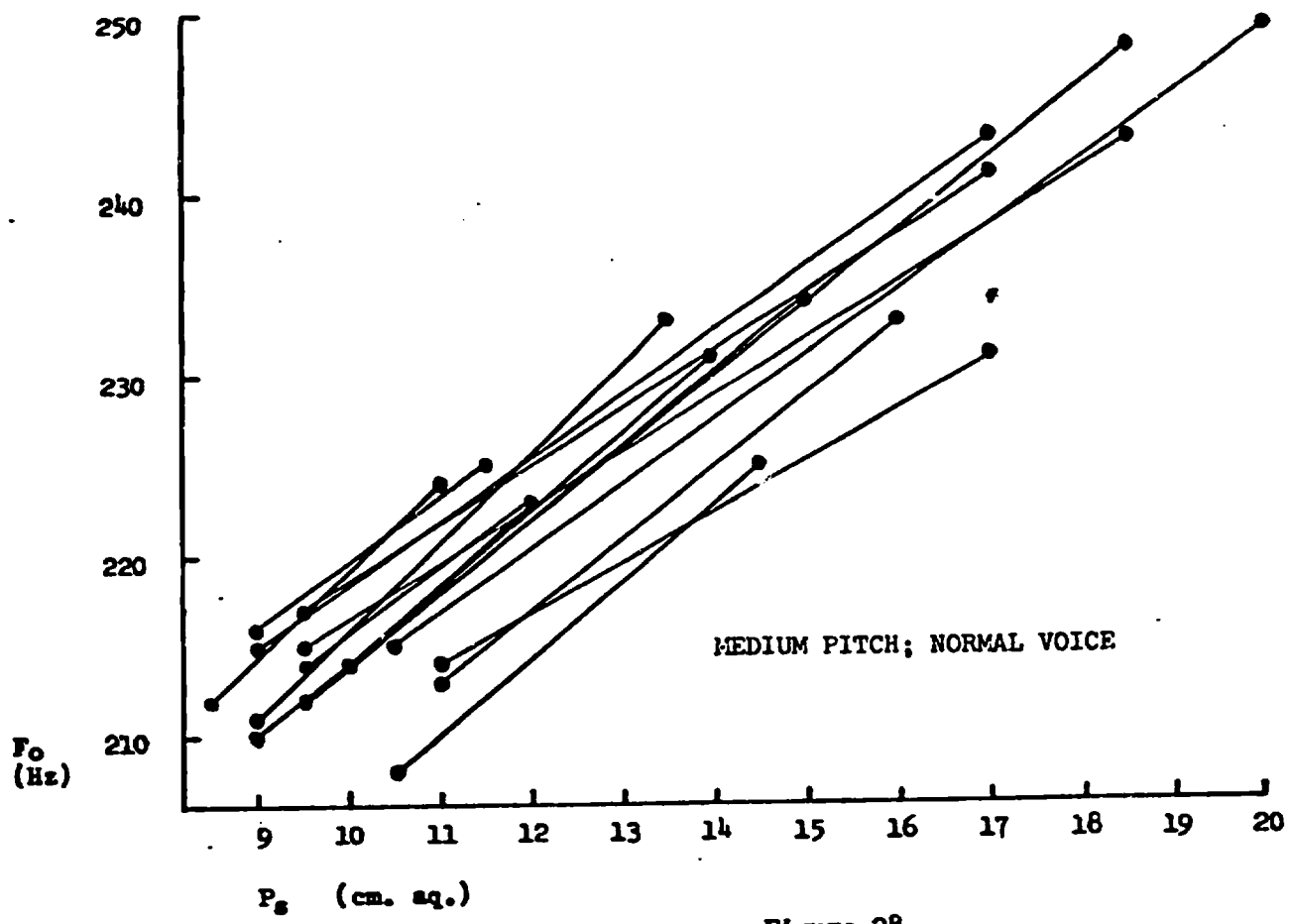

250
240
230
220
210
F_o (Hz)
9
10
11
12
13
14
15
16
17
18
19
20
P_s (cm. aq.)
MEDIUM PITCH; NORMAL VOICE

Figure 28.

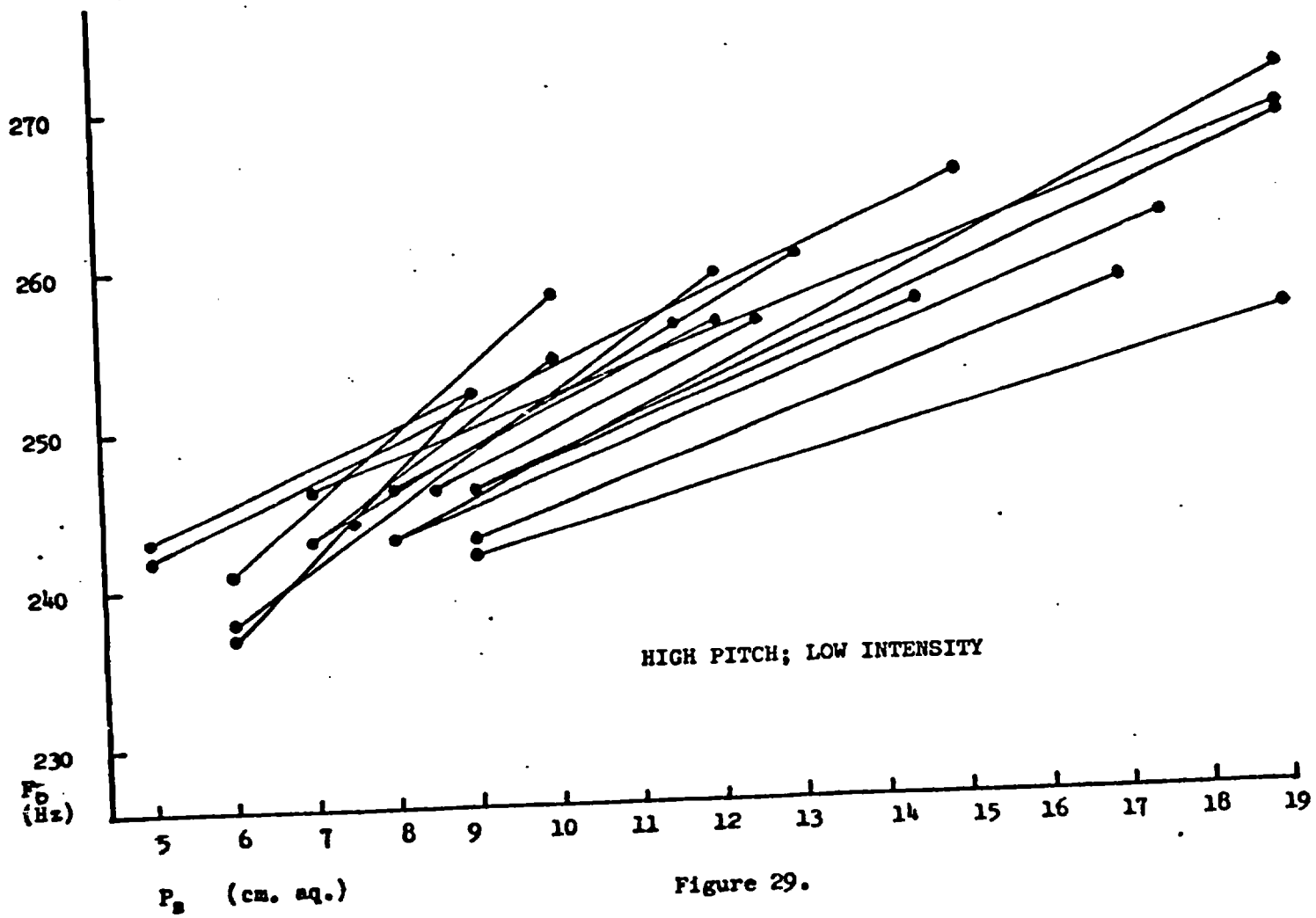

270
260
250
240
230
F_0 (Hz)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
P_s (cm. aq.)
HIGH PITCH; LOW INTENSITY

Figure 29.

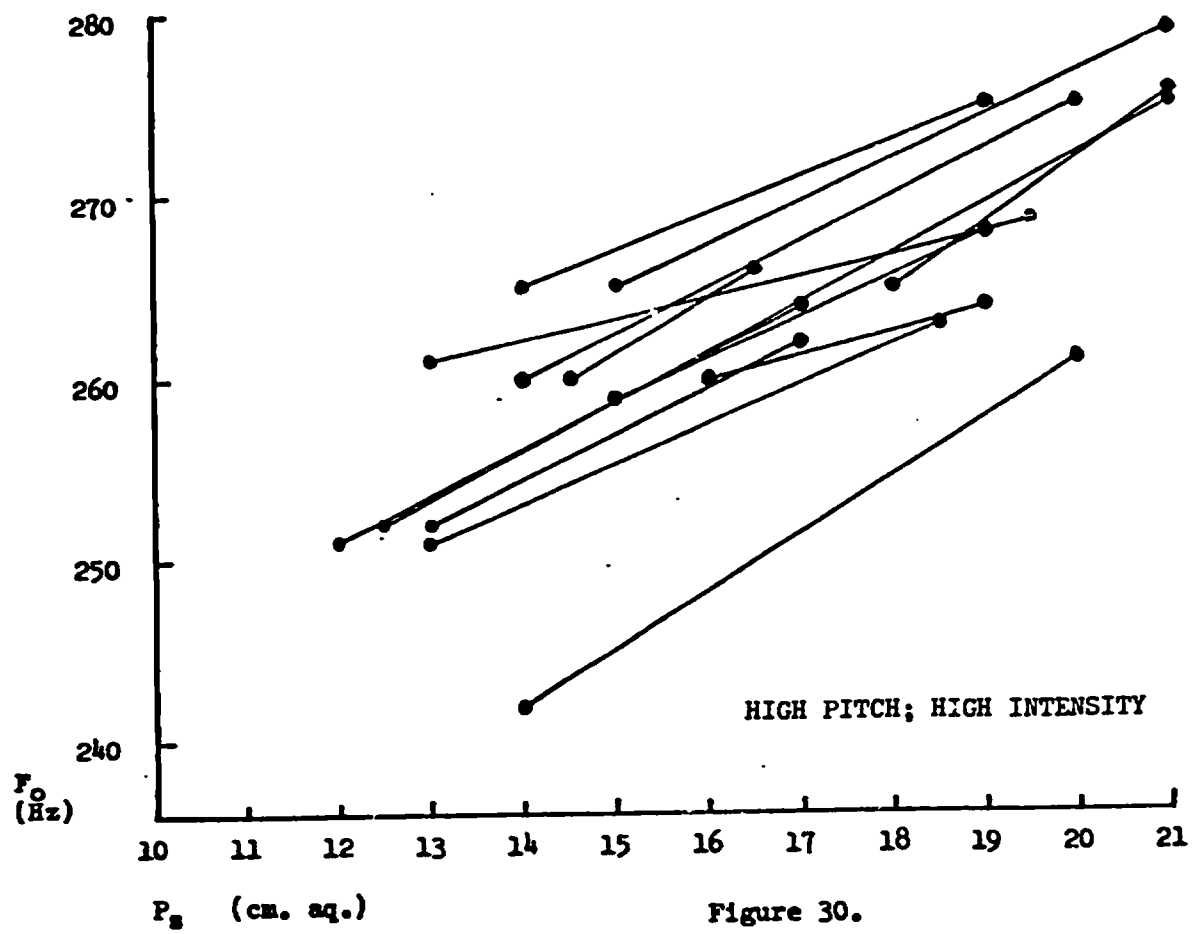

280
270
260
250
240
F_o (Hz)
10
11
12
13
14
15
16
17
18
19
20
21
P_s (cm. aq.)
HIGH PITCH; HIGH INTENSITY

Figure 30.

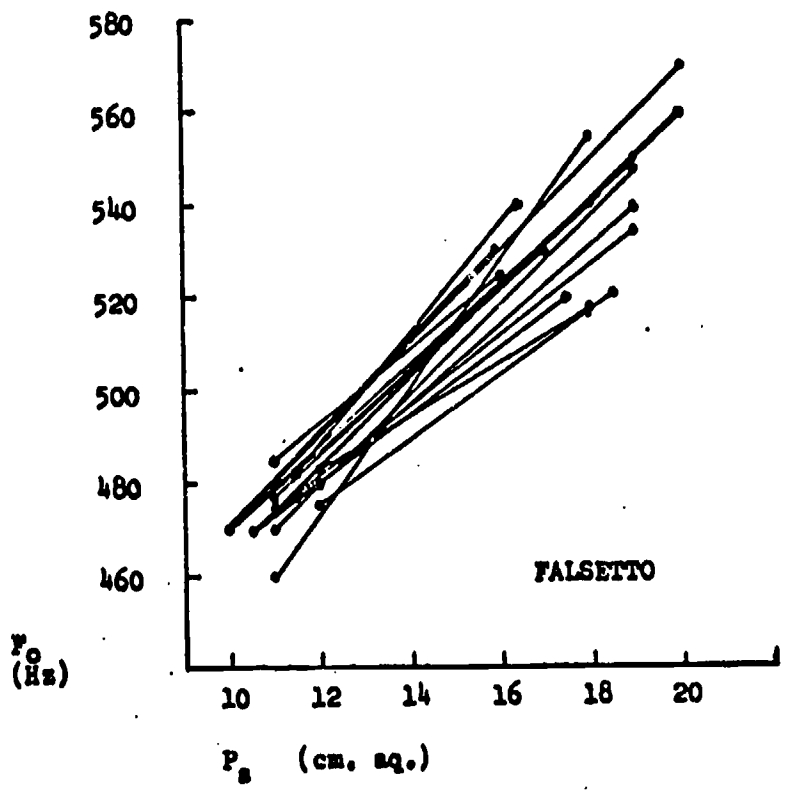

580
560
540
520
500
480
460
F_o
(Hz)
FALSETTO
10
12
14
16
18
20
P_s (cm. aq.)

Figure 31.

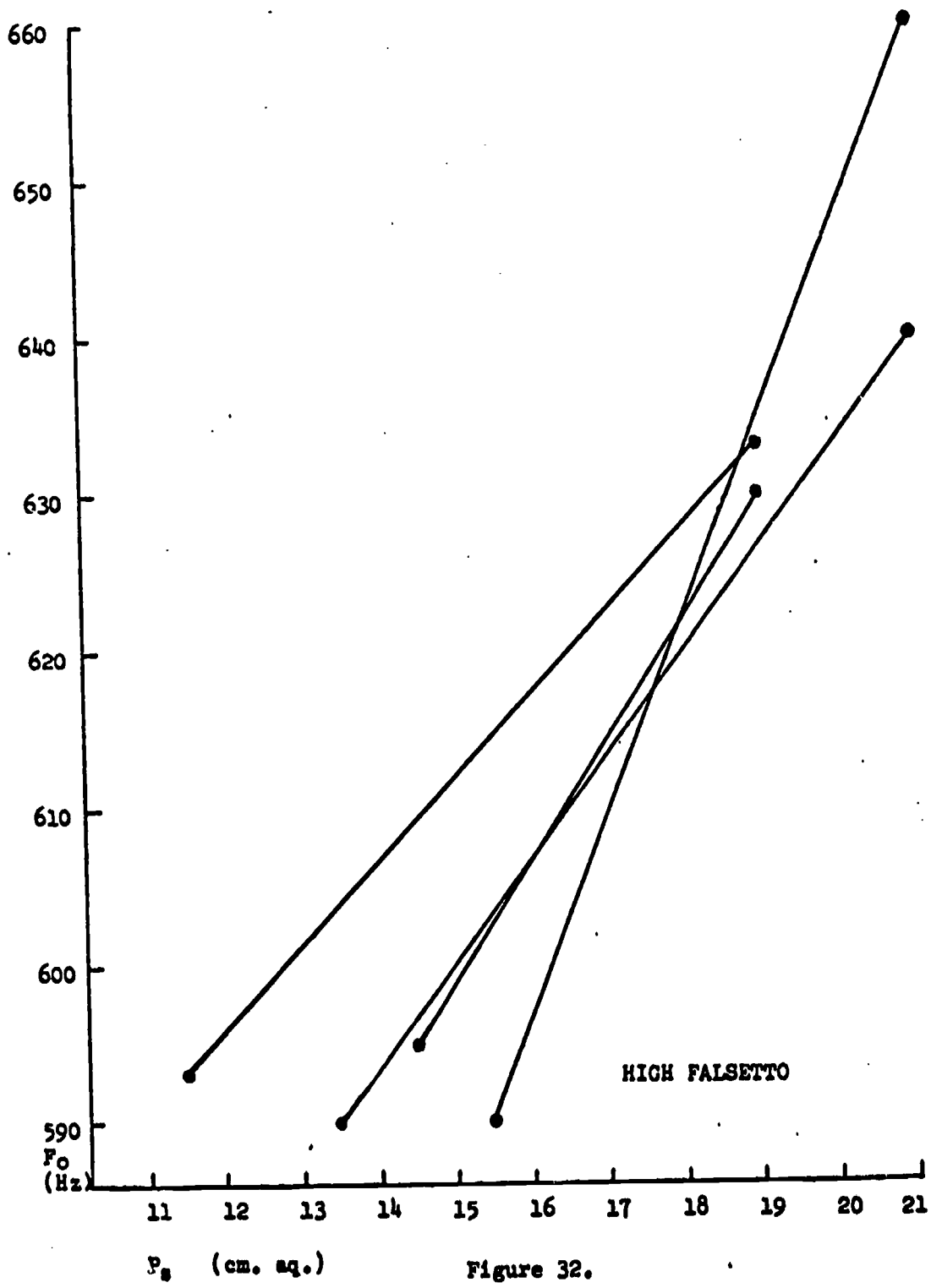

660
650
640
630
620
610
600
590
Fo
(Hz)
11
12
13
14
15
16
17
18
19
20
21
Ps (cm. aq.)
HIGH FALSETTO

Figure 32.

kind reported above, and in fact made no attempt either to control or to monitor the highly relevant variable of laryngeal adjustment. He merely assumed that the laryngeal tension was constant. As has been shown above this is not a valid assumption.

This study obviously needs improvements, not the least of which would be to include a wider variety of conditions of phonation including using different vowels. Also it shows the need to administer the pushes rapidly and randomly, perhaps by some kind of machine which was controlled by a computer emitting triggering pulses at random intervals. Furthermore, it should be established whether the change in the cricothyroid muscle that was encountered on some pushes was due to predictive behavior on the part of the subject or due to some laryngeal reflex triggered by greater air flow through the glottis. And it should be determined whether this effect can be completely eliminated either by randomizing the pushes or by anesthetizing the mucosa of the larynx.

Until these improvements are accomplished, however, the procedures used here would appear to constitute an approximate but valid calibration of the effect of subglottal pressure on fundamental frequency.

STUDY THREE: EMG PLUS SUBGLOTTAL PRESSURE

The third study in this series consisted of sampling simultaneously in one subject, (JO), the following parameters: the acoustic signal (from which the F_o was extracted), the subglottal pressure, and the muscle action potentials of the cricothyroid and the lateral cricoarytenoid. The techniques for obtaining these parameters and the basic text of utterances used were the same as in the previous studies, although extra sentences were also included. The purpose was to find out if the two factors which could affect the F_o, pressure and larynx adjustment, worked together or not, and which of the two, if either, proved to be more closely synchronized with the corresponding increases in F_o.

Results and Discussion

Figures 33 through 38 present samples of the resulting data. Only the pitch curve has been slightly retouched to remove artefacts and to make it agree more with the pitch curve derived from narrow band spectrograms. Not surprisingly it was found that increases in both subglottal pressure and laryngeal adjustment often occurred at about the same time on accented syllables. This is shown in Figures 33b, 34a, 35b, 36a, and 36b. Our subglottal pressure records do not differ in any gross way from those of the other investigators cited above. At the start of phonation the pressure rises rapidly as the vocal cords close; at the end it falls as the vocal cords open and the

Figure 33a. Subject: JO; utterance: "Bev bombed Bob," unprocessed signals, from top to bottom: fundamental frequency, subglottal pressure, EMG of cricothyroid muscle, EMG of lateral cricoarytenoid muscle, audio, and time standard (0.1 sec.).

Figure 33b. As in Figure 33a; utterance: "BEV bombed Bob."

Figure 34a. As in Figure 33a; utterance: "Bev BOMBED Bob."

Figure 34b. As in Figure 33a; utterance: "Bev bombed BOB."

Figure 35a. As in Figure 33a; utterance: "Joe ate his soup."

Figure 35b. As in Figure 33a; utterance: "JOE ate his soup."

Figure 36a. As in Figure 33a; utterance: "Joe ate HIS soup."

Figure 36b. As in Figure 33a; utterance: "Joe ate his SOUP."

Figure 37a. As in Figure 33a; utterance: "Bev bombed Bob?"

Figure 37b. As in Figure 33a; utterance: "Did Bev bomb Bob?"

Figure 38a. As in Figure 33a; utterance: "Did BEV bomb Bob?"

Figure 38b. As in Figure 33a; utterance: "Did JOE eat his soup?"

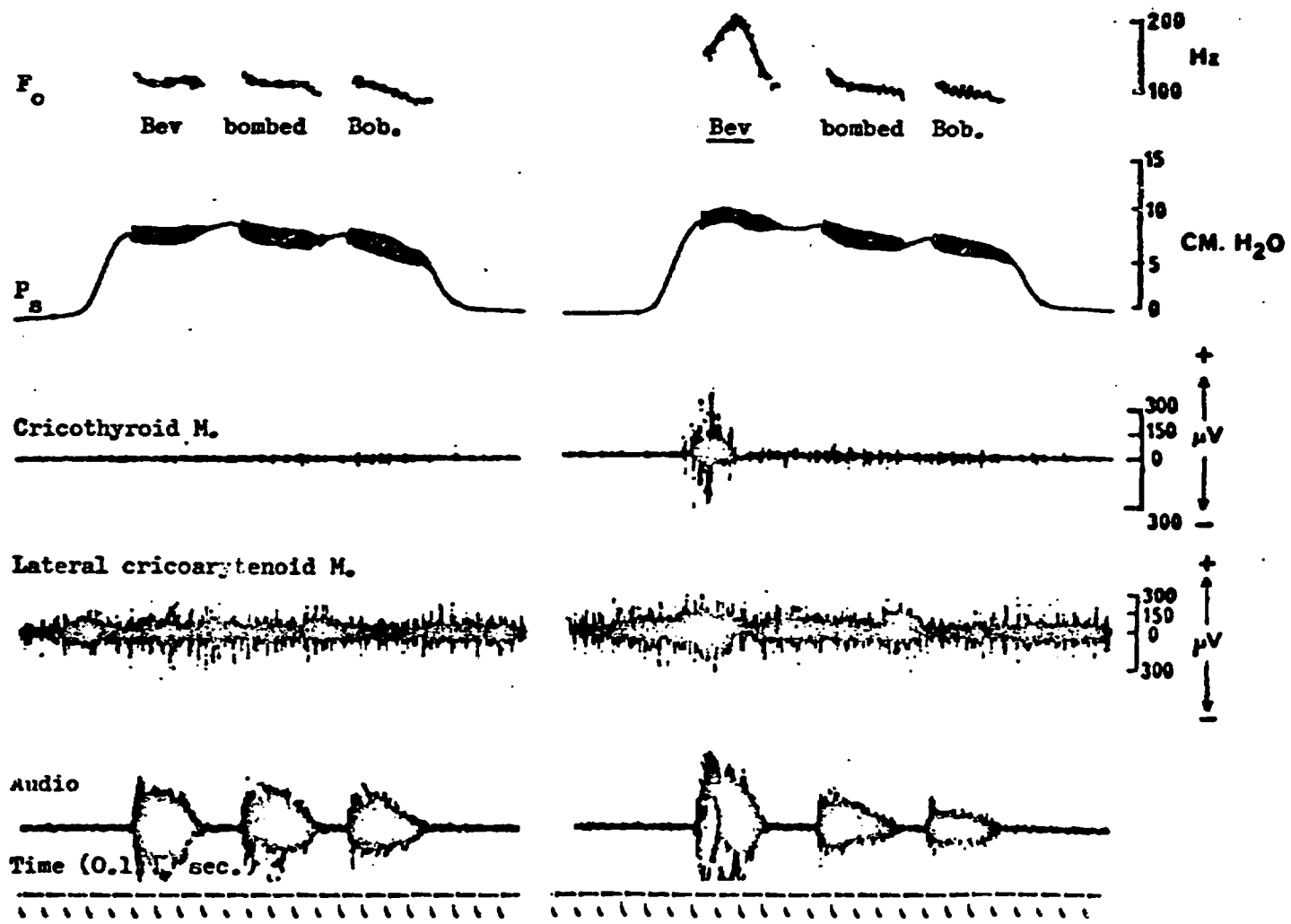


Figure 33a. Figure 33b.

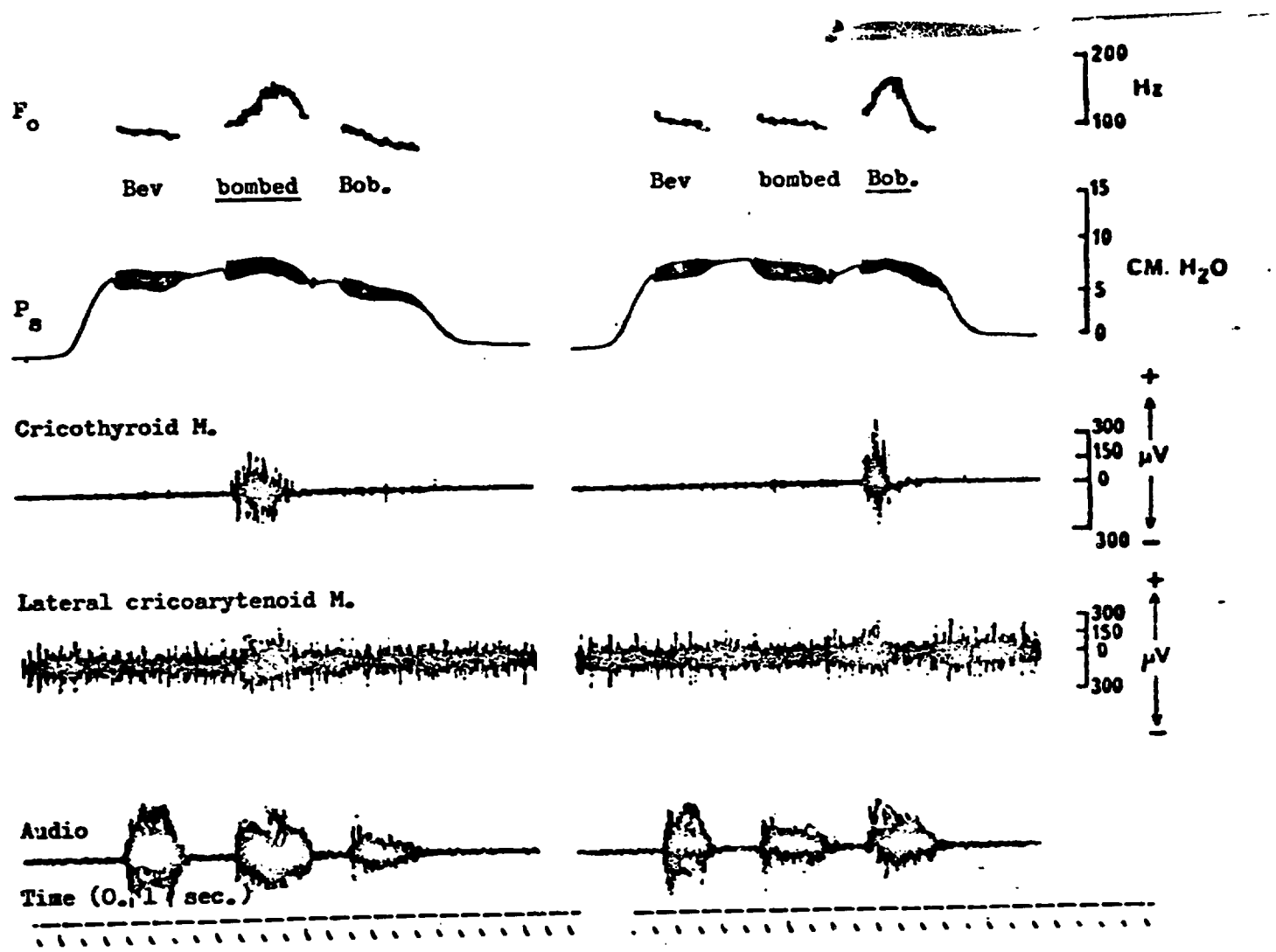


Figure 34a.

Figure 34b.

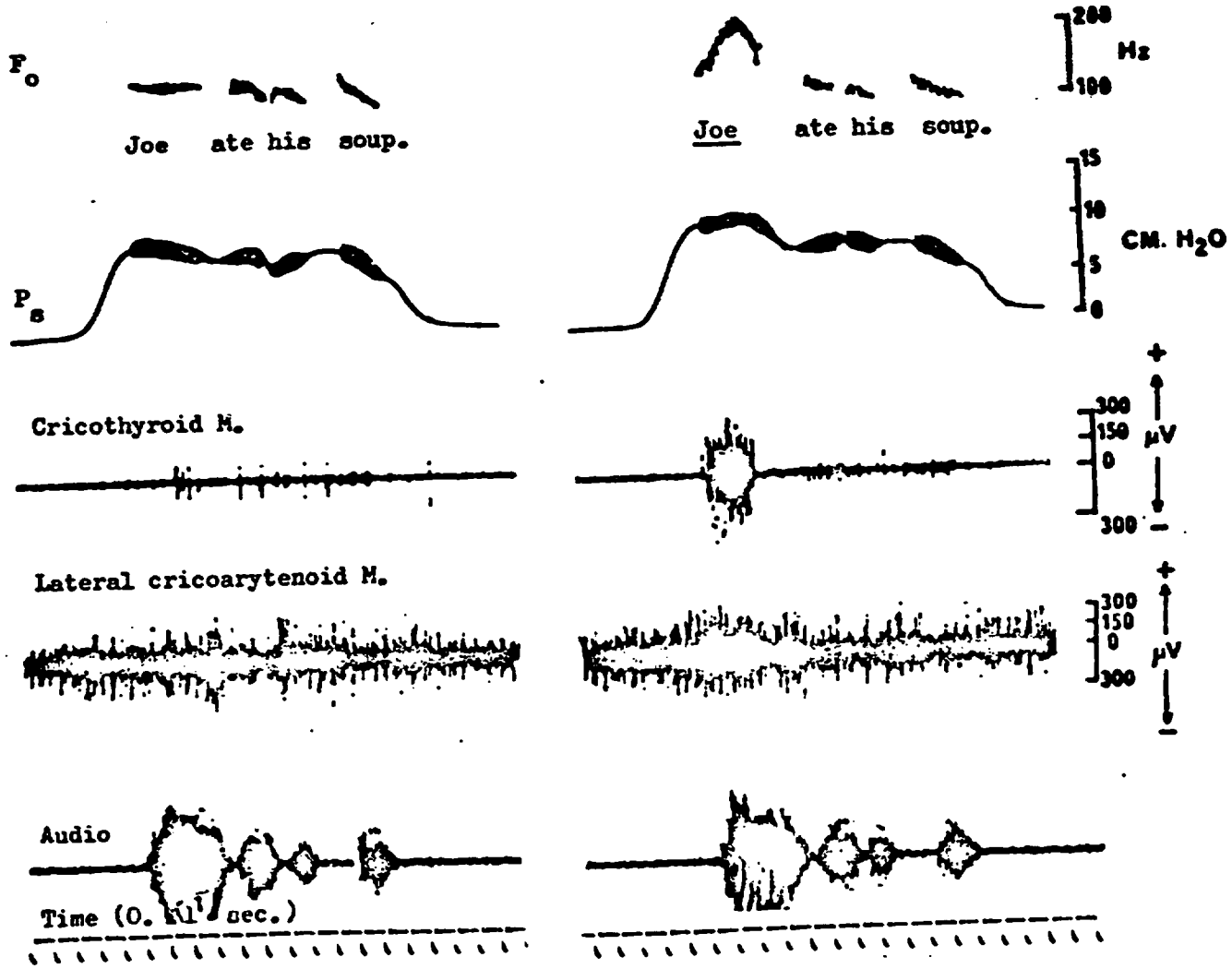


Figure 35a.

Figure 35b.

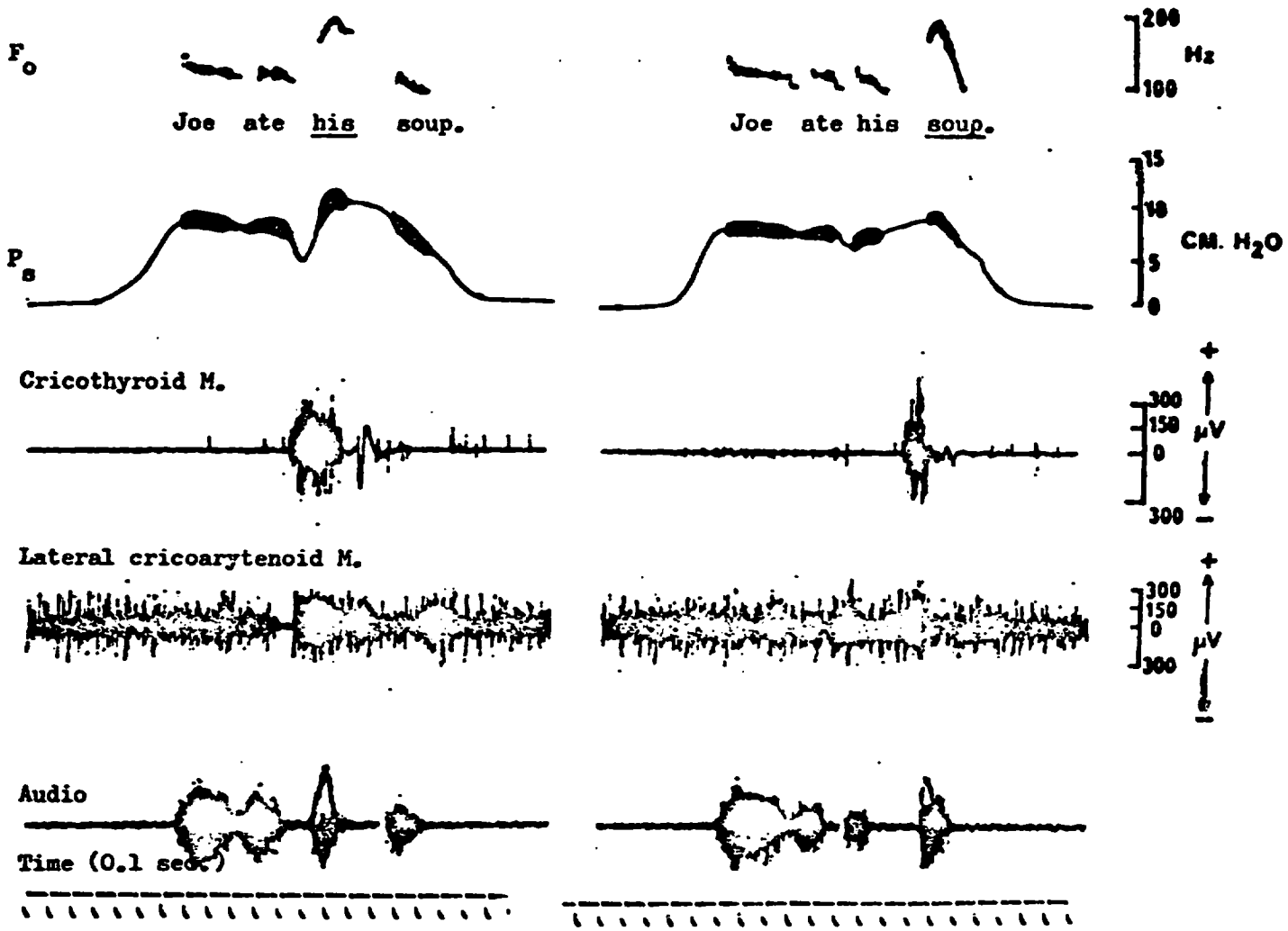


Figure 36a.

Figure 36b.

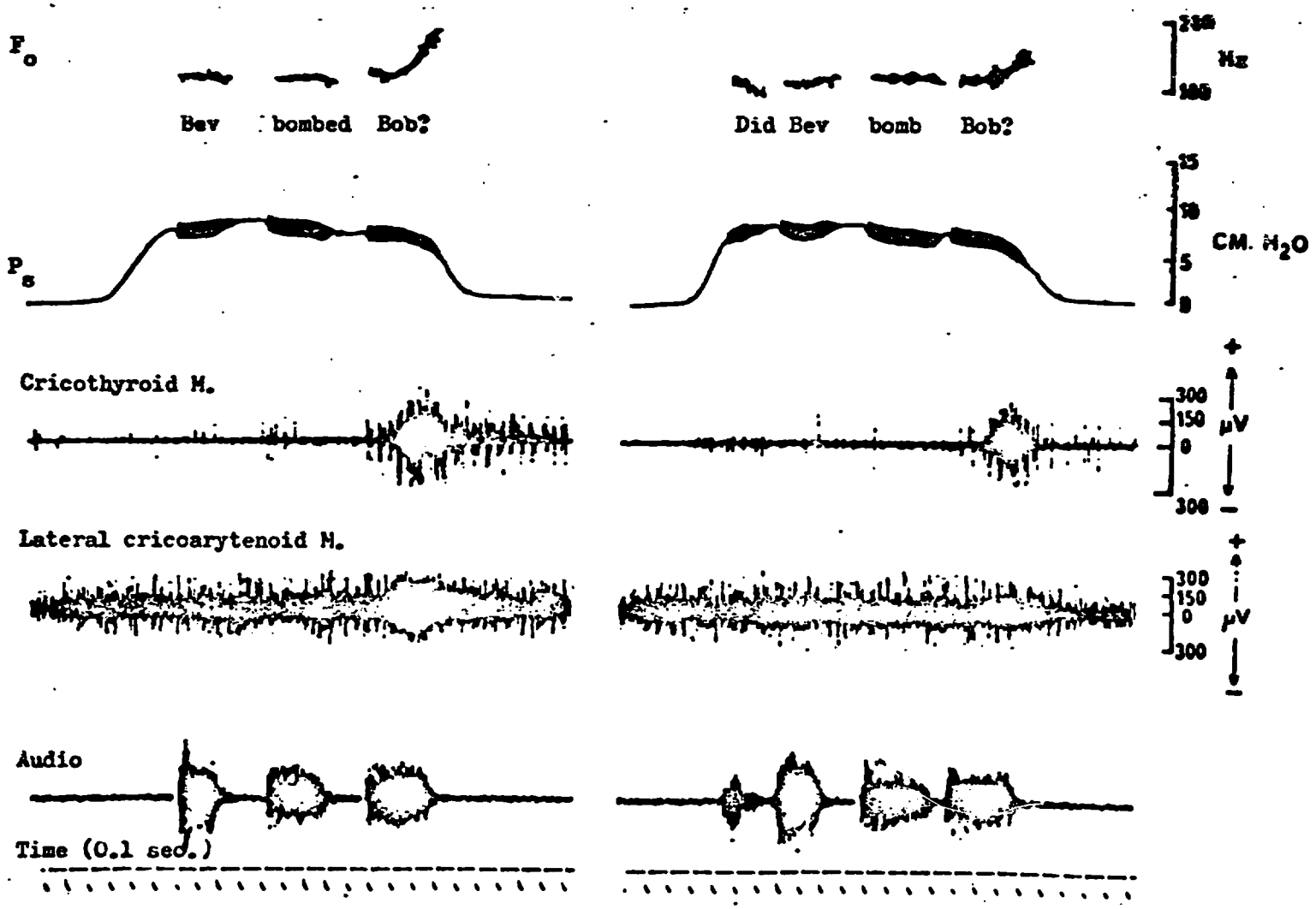


Figure 37a.

Figure 37b.

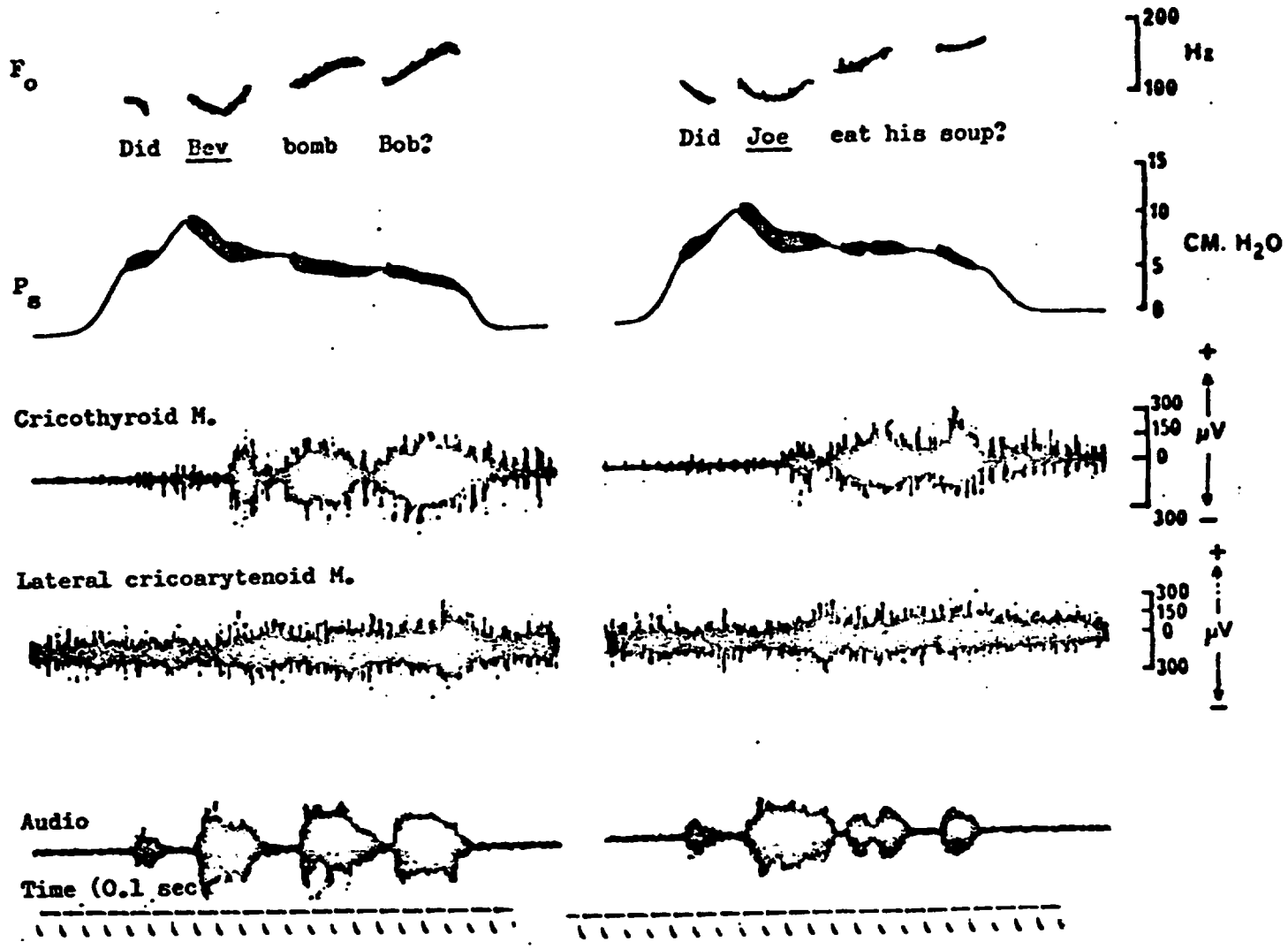


Figure 38a. Figure 38b.

respiratory cycle approaches inspiration; in between it maintains a fairly steady positive level. Momentary increases are found on "stressed" or accented syllables and momentary decreases are found when the glottal resistance drops as it does during true /h/'s, cf. Figure 36a and the drop in P_s during the /h/ in "his."* Activity of the vocalis muscle was not recorded in this study but it is known from records of this muscle's activity during the production of similar sentences that it typically shows increased activity on the emphasized words. Thus we might reasonably expect that glottal resistance increases during these words. This could account for part of the observed subglottal pressure peaks.

It is interesting to note that in all cases except in Figures 38a and 38b the subglottal pressure peaks occur during the vocalic portion of the emphasized words. In the mentioned figures, however, the rather large pressure peaks occur during the initial stops of the words "Bev" and "Joe," respectively. It is not clear why the peculiar intonation contour used (in Bolinger's system, Accent "C") should cause this, but it does suggest that the observed pressure peaks are due in part to the increased supraglottal resistance created by the consonant closures.

The EMG records are also similar to those of the first study and again show these two intrinsic laryngeal muscles contracting during and just preceding a pitch rise. It is easy to see how, if one sampled only the pressure, as Lieberman did, and inspected it superficially, one could arrive at the erroneous conclusion that the F_o was determined solely or primarily by the pressure. The correlation between pressure

---

* In Figure 35b, however, the sentence was pronounced as [dʒoʷɡɪɾɪsuʷp], i.e., without a /t/ or an /h/ in "ate his" and consequently there is no drop in the pressure trace. Phonetic reduction of such words in unaccented positions is quite common and occurs as well in Lieberman's data in his Figure 3 (1968b, p. 33; which is the same as Figure 4.10 in Lieberman 1967a, p. 67). That is, it is clear from the spectrographic record in that figure that the speaker had a voiced flap and did not pronounce a /t/ or /h/ in the words "ate his." Thus it is puzzling to find Lieberman commenting on this figure as follows:

> Note that the fundamental frequency function in FIGURE 3 is quite smooth through the words "ate" and "his" despite the fact that phonation was interrupted when the speaker abducted his vocal cords to produce the consonants /t/ and /h/ in these words. [1968b, p. 34]

and F_0 at first glance seems quite good and in this study, too, one could have been plotted against the other and the slope of the data points taken as the value $\Delta F_0/\Delta P_s$. But this would be quite misleading since in this case the laryngeal muscles were clearly participating in the regulation of F_0.

A more careful examination of the data, even without the EMG records, would prevent one from jumping to the questionable conclusion that the F_0 is determined primarily by the subglottal pressure. It is true of our data -- as it apparently was of the data of Öhman and Lindqvist -- that the pressure rises are not always "in phase" with the F_0 rises. This is evident in Figures 33b and 34a, of "Bev bombed Bob," and "Bev bombed Bob," respectively. In other cases, where the peak pressure and peak pitch are roughly simultaneous, the rest of the curve frequently doesn't correlate well: whereas the pressure is usually increased over the entire emphasized syllable by a fairly constant amount (c. 1 or 2 cm. aq. relative to the same sentence without emphasis) the F_0 is usually rapidly changing at the same time, cf. Figure 35b of "Joe ate his soup." Other discrepancies between F_0 and pressure can be found. In fact this is true of Lieberman's (1967a) data, too, as has been pointed out by Vanderslice (1967). Although Lieberman allows for a ± 40 msec. temporal uncertainty due to synchronization problems, the F_0 peaks and pressure peaks show lack of synchronization by approximately 130, 75, 150, and 60 msec. in his Figures 4.14, 4.19, 4.24, and 4.26, respectively.

A qualitative estimate of the extent of contribution of these two parameters can be gained by comparing such pairs of sentences in our data as Figures 33b and 38a, 35b and 38b. Without the assistance of the laryngeal muscles in Figures 38a and 38b the relatively high subglottal pressure produces little or no rise in F_0; however when the laryngeal muscles do help, as in Figures 33b and 35b the pitch rise is considerable.

GENERAL DISCUSSION

The experimental studies reported need various improvements as noted. Even so, on the basis of these findings and on the literature reviewed, particularly the clinical studies relating to pitch control, it seems safe to conclude, with Sweet (1877, p. 93), Ladefoged (1964, p. 41), Fry (1964, p. 217-218), Öhman and Lindqvist (1966 a, p. 4), Öhman (1967b, p. 29), Vanderslice (1967, p. 76) and Proctor (1968, p. 208) that the pitch in speech is mainly determined by the action of the laryngeal muscles. There must be a certain minimum subglottal pressure, of the order of 2 or 3 cm. aq. (and, of course, approximation of the vocal cords), in order for phonation to be

sustained, but beyond that, variation in the pressure seems to have relatively little effect on the pitch, but, of course, affects very much the intensity of phonation (Ladefoged, 1963). This holds true at all places in the utterance; at the end, the beginning or the middle, and independently of the grammatical structure of the utterance. This conclusion permits us to apply confidently to speech the wealth of quantitative data already accumulated on the action of the larynx in non-linguistic phonation, with one point of caution, namely to keep separate the larynx's participation in prosodic and segmental gestures (e.g., glottal stop, voicelessness). And the fact that new techniques have been developed for investigation of laryngeal parameters during speech, -- e.g., photo-electric glottography (Malécot and Peebles 1965, Ohala 1966, Lisker, Abramson, Cooper and Schvey 1966), fibre-optic laryngoscopy (Sawashima and Hirose 1968), transcutaneous electromyography of the laryngeal muscles (Hirano and Ohala 1969), and ultrasonic scanning (Minifie, Kelsey, and Hixon 1968), suggests that in the near future more of the much-needed investigation of "segmental" laryngeal activity will be done, along with a continuation and refinement of the study on the physiology of intonation.

Where does this leave the situation with respect to linguistic theories of intonation? It leaves it back where it was before Lieberman's innovative theory. It was at least implicit in all traditional theories of intonation that a human speaker can and does "program" his laryngeal muscles to execute any pitch change whatever, limited only by the intrinsic mechnical and neuromuscular constraints of the larynx and its muscular and cartilaginous attachments. Which contour the speaker does execute is not wholly predictable from the lexical content or syntactic structure of the sentence -- at least not from such a limited range of syntactic structure such as whether or not the sentence is a yes-no question, WH-question, non-final or final phrase. This has long been known to the traditional writers on intonation such as Palmer and Blandford (1924), Armstrong and Ward (1926) and Pike (1945). Everyday evidence for this point is offered by the fact that sentences which are structurally identical from a lexical and syntactic point of view can have different intonation patterns:

Have you been to Disneyland? or Have you been to Disneyland?

The different intonation contours are determined by the speaker's attitude, or, if one prefers more current terminology, by a difference in the deep (very deep) structure of the sentence. Actually, including such information, that is, the speaker's attitude towards what he says, has probably ceased to be a stumbling block for current linguistic

theory, now that the boundary between syntax and semantics is rapidly vanishing (cf. McCawley 1968; Menzel forthcoming; Heny, forthcoming).

If Lieberman's theory of intonation is rejected does it leave us with no alternative theory of comparable scope? Yes and no. It may be that there is no single theory that can explain the intonational phenomena in English and all other languages of the world, plus the characteristics of infant cry and the results of experiments such as that by Hadding-Koch and Studdert-Kennedy. The history of science has seen other theories, such as the ones offered by phrenology and Freudian psychology of the last century, which, once rejected, have even now never been replaced by any theory covering the same wide range of phenomena. But although this work does not offer a complete alternative theory of English intonation, others do, of course, exist. Such works on English intonation as those by Palmer (1922), Palmer and Blandford (1924), Pike (1945), Bolinger (1958) and others have certainly covered or catalogued a wider range of intonational phenomena than was treated in Lieberman's theory, and they treated many points more accurately. They knew, for example, that accent or emphasis on words could be manifested by drops in pitch as well as rises in pitch. Moreover, such recent intonational theories as those of Vanderslice (1968a and 1968b) for English and Öhman (1968) and Öhman and Lindqvist (1966a) for Swedish incorporate many of the valuable points made by the earlier studies but go well beyond them by turning to the task of attempting to synthesize correct intonation. They begin with posited linguistic units and through a system of rules generate the continuously varying F_0 accompanying sentences. Both of these theories rest on the assumption — which I believe the data presented here has shown to be correct — that the speaker can execute any pitch pattern he wants to by using his laryngeal muscles. The field does have adequate alternatives to Lieberman's theory of intonation.

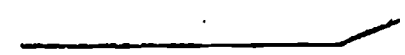
There are a few other minor points of Lieberman's theory which deserve comment. Lieberman's notions on "least effort" and "simplicity" lead him to make two other claims which the facts — including his own — do not support. The first is:

> The "archetypal" articulatory correlate of [+P_S] is a momentary increase in the subglottal air pressure that can occur in any part of the breath group except the last 150-200 msec. of phonation. [1967a, p. 105]

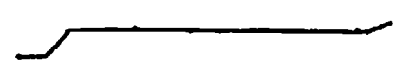
Presumably this is because the last 150-200 msec. are reserved for manifesting [+BG] or [-BG], i.e., the terminal pitch contour. For whatever reason, the principle is contradicted by Lieberman's own data: his Figures 4.12, 4.15, 4.20, 4.27 all contain momentary pressure peaks that are within 150 msec. or less of the end of phonation. The second claim is:

> ... it is unlikely that any language has a laryngeal tension function that rises before the terminal part of the breath-groups, since this could lead to confusion between the marked and the unmarked breath-groups. [1967a, p. 105]

This claim is falsified by the results given above from EMG studies of the intrinsic laryngeal muscles, but we need look no further than Lieberman's own data. Pitch rises occur well before the end of the breath-group in his Figures 4.13, 4.14, 4.15, 4.22, and 4.24, all of which are various renditions of "Did Joe eat his soup?" and all of which have a clear pitch rise on "Joe." With the remotely possible exception of 4.24, none of these sentences have a pressure curve that could be construed as causing the pitch rises. The only alternative is the laryngeal tension. A point quite clearly missed by Lieberman is that the "early rise" (Pike's term) in such sentences is a well-known phenomenon in English intonation (Palmer and Blandford, p. 14-17; Pike, p. 75). Katz and Postal more recently (1964) have commented on this (p. 106). To oversimplify, the principle is that an emphasized word in a sentence draws the intonation contour to it whether it is at the end of the sentence or not. Thus:

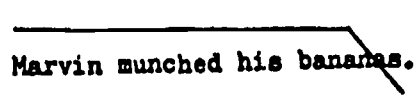


Marvin munched his bananas?

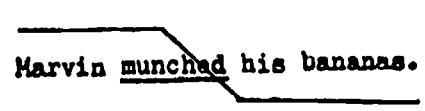


<u>Marvin</u> munched his bananas?

It works as well for falling contours:



Marvin munched his bananas.



Marvin <u>munched</u> his bananas.

In fact, knowing this principle, one is able to judge that the sentence in Lieberman's Figures 4.13, 4.15, 4.22 and 4.31 were misread by his subjects: they placed emphasis on "Joe" when it was not called for.

Neither Lieberman's data nor the data of the present study provide any confirmation or justification for two other concepts used

by Lieberman. His use of the terms and concepts of "marked-unmarked" -- unmarked for the breath-groups ending in a falling pitch and marked for the one ending in a rising pitch -- rested on the assumption that to raise the pitch at the end required more effort (some *extra* action) than to merely allow the pitch to fall naturally at the end due to the falling subglottal pressure. In fact muscular activity is involved in both actions. One could even argue that if we had to assign the labels marked-unmarked to these two contours, the unmarked member ought to be the one ending in pitch rise, since more and larger muscles are apparently involved in lowering pitch (the extrinsic laryngeal muscles) than are involved in raising pitch (cricothyroid primarily, plus vocalis and lateral cricoarytenoid). However we do not believe that the concept of marked-unmarked is useful in this case and would recommend that the terms not be used at all, or at least not until more research is done which would justify the assignment of such terms. We must guard against the universal tendency in science of thinking that we have somehow solved or explained a problem by re-labelling it.

Another concept for which no evidence can be found is that of "archetypal physiological correlates" versus "alternate articulatory maneuvers." It is necessary at the outset to dispose of the notion -- which Lieberman explicitly espouses -- that occurrences of non-archetypal gestures do not contradict his theory. If every case that doesn't fit the theory can be explained away as an "alternate articulatory maneuver," then in fact there is no theory at all because there is no potential for falsification. Thus a theory that claimed "all cats are black, except those that aren't" is uninteresting because there is no universe in which it could be false. Such theories command no scientist's or anyone else's attention. Thus we must pay attention to " exceptions" -- and insist that to be exceptional they can only occur occasionally (Vanderslice 1967). From the data presented above it appears that what Lieberman calls "alternate articulatory maneuvers" are overwhelmingly the rule and not the exception. But this in itself does not falsify the rather more important content of Lieberman's theory, namely, that, first, the physiological and acoustic correlates of [-BG] are indeed "archetypal," that is, "innate" (since the breath-group is included in the behavioral repertory of the newborn infant) and are thus universal, and that, second, the features [±BG] and [+P_s] are psychologically real. These claims could, theoretically, be true in spite of the known physiological facts. If the first claim were true it would be of major importance because, first, it would be about the only really convincing physiological explanation for the origin of any one aspect of speech (namely the falling fundamental frequency at the end of sentences) and second, it would provide a solid basis for explaining the universal presence of this pitch contour in all human languages. If the second claim were true it would be a significant fact for theories

of speech perception. It is important, then, that the evidence for these claims be examined carefully. This is done in the next two sections.

Examination of Evidence From Infant Cry

The assertion made by Bosma, Truby and Lind (1965) that the innate neuromuscular "programs" which enable the newborn infant to cry form the basis, during maturation, of certain aspects of speech, is eminently sensible and not very surprising. We could as well note that the innate ability of the infant to thrash his arms and legs in the air forms, later on, the basis, in some sense, of his being able to walk, run, and throw a ball. Although the nature of the built-in "programs" may be much more detailed than we now suspect, there is every reason to believe that the subsequent neural programs enabling such mature skills as talking and walking are several orders of magnitude more detailed and complex than these for simple crying and arm thrashing. To use a computer analogy (as in fashionable these days), every computer comes with some built-in ("innate") interpretive programs or logic circuits which enable it to perform a particular set of commands given it by the user. However, the ability of the computer to perform "add" 's and "complement" 's does not begin to compare with its acquired ability via a program consisting of many "add" 's, "complement" 's and many other commands to guide a spacecraft to the moon. It is thus reasonable to assume that by studying in detail the behavior of newborn babies we might indeed gain some clues as to the kinds of elementary gestures or programs they are endowed with in the womb, which in combination and when more precisely controlled become the basis for their later more complex motor skills. This is not a new idea (Darwin 1872, Lorenz 1958, Andrews 1965) even in phonetics, but in so far as it has not been current in phonetics, Lieberman has made a significant contribution to the field by re-focussing attention on a neglected but promising source of evidence on the neuromuscular substrate of speech.

But Lieberman goes beyond this, claiming that there is such a striking similarity between certain aspects of infants' cries and adults' intonation contours as to permit the conclusion that the latter are based on, and derived from, the former -- i.e., the innate pattern is taken over into speech essentially unmodified except that other patterns may be superimposed on it. Although many different types of infant vocalization can be differentiated, Lieberman suggests that one important dichotomy exists: namely, attention-getting cries versus babblings or sounds which do not seek to attract attention. The former are labeled "innate referential breath-groups" and are said to be loud, shrill, and have a rising-falling F_0 contour with an abrupt

drop in F_o at the end. The other sounds and vocalizations of the infants are said to be softer, less shrill and not always ending with falling pitch.

Lieberman (1967a) focuses attention primarily on two points: (a) the variations in F_o at the end of these "innate referential breath-groups", specifically, asserting that it rapidly falls at the end, and (b) on the correlation between the subglottal air pressure and F_o throughout the entire breath group. The only source of information he refers the reader concerning the extent of correlation between subglottal air pressure and F_o during infant cries is the research of Bosma, Truby, and Lind.* They actually recorded esophageal pressure which, however, can give a rough indication of subglottal pressure variations. Lieberman's report of their published findings is highly misleading, however, in several ways. The esophageal pressure recordings during cries were obtained from either 5 (p. 64) or 6 (p. 11 and p. 66) infants, not 30, as is implied by Lieberman. This false impression of the sample size could easily be obtained if one read only the brief report of their work in Bosma (1964) and not the more extensive report in Lind (1965) -- the cries of 30 infants were subjected to acoustic analysis, though. Further, Bosma (1964) and Bosma, Truby, and Lind (1965) publish the esophageal pressure measurements and spectrographic displays of only 13 cries of one infant. Of these, five show the pressure recording going off scale in the middle of the cry. In Truby and Lind (1965) two other pressure curves from the same infant are published, the F_o of which can be obtained by analyzing the corresponding cries on the phonograph record accompanying the book. Thus the number of cries on which Lieberman bases his conclusion of the extent of gross correlation of subglottal pressure and F_o (unless he has access to other data which has not been published) is about ten, all produced by the same infant. Lieberman reports

> The measured esophageal pressure functions for the cries all had a similar shape. The esophageal pressure gradually rose from the start of phonation to either a level or a slightly falling "plateau." The esophageal pressure then abruptly fell prior to inspiration ... The "shape" of the fundamental frequency contours of the cries was similar to the shape of the typical esophageal pressure contour. Qualitatively speaking, *the gross variations of the funda-*

---

* The reference to their work is cited incorrectly by Lieberman. What is given as Bosma, Lind, and Truby (1964) should be Bosma (1964) — where the work of the three authors is summarized in a brief report, being put forth in much more detail in four long articles in the book *Newborn Infant Cry* (Lind 1965).

mental frequency contour thus seem to be a function of the subglottal air pressure during infant cries. The fundamental frequency, which rises initially as the subglottal air pressure builds up, remains level or slowly falls until the end of expiration, when it abruptly falls. [p. 43 — italics mine]

The sentence italicized above might be justified if every other point in the quotation were true, but a casual inspection of the published data shows that many of them are not. That the pressure recordings for this single infant all have a similar shape is true. Bosma remarks on this point as well. However Bosma's point was precisely that many of the physiological patterns of cries are characteristic of the individual infant. This is certainly quite different from asserting that the esophageal pressure curves were similar for all of the infant subjects. Secondly, during the cries the esophageal pressure curves do rise from the start of phonation, but after this initial rise -- which, by the way, occupies the greater part of any cry -- a falling "plateau" can be seen in only two cries (the extra two published in Truby and Lind), a short level plateau could be claimed in four other cries, and in the remaining four there is nothing even faintly resembling what might be called a plateau (i.e., a "flat portion"). At the end of each cry the pressure does indeed fall abruptly, as, of course, it must if the infant is to get a new supply of air into his lungs. Finally, it is not clear whether the assertion that the F_o contours of the cries resembled the pressure curves refers to the ten cries with accompanying pressure curves or to all the hundreds of cries studied by Bosma and associates. In either case the assertion is untrue. In five of the ten cries a two-octave upward "shift" occurs in the middle. Prior to this shift the F_o is relatively steady or even falling while the esophageal pressure is rising. In the other cries there are variations in the F_o which do not consistently follow the corresponding variations in the pressure. The statement applies no better to the entire collection of cries. First, as noted by Bosma (1964) and by Truby and Lind (1965), infants' cries are highly individual and often incorporate various kinds of modifications of what is termed a "basic cry," among which modifications are octave (or more) shifts in F_o, addition of turbulence, introduction of sudden silence in the cry, vibrato, etc. Second, setting aside all these non-basic cries, the "basic" cry of infants has an almost symmetrical rise-fall F_o contour, but the pressure curves show no such symmetry. That the F_o of the cries abruptly falls at the end was not a stated conclusion of Bosma and his colleagues, and in the acoustic records of cries published in their work, one can find about an equal number of cries which do and do not exhibit this sudden downward movement of F_o at the end. There are even a few cases where the F_o abruptly rises at the end.

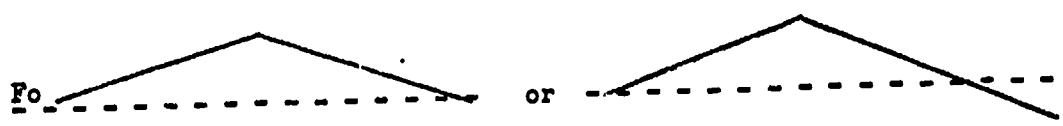
Lieberman also refers the reader to the work of Ostwald (1963) to support the notion that the F_o of cries falls abruptly at the end. Reporting on Ostwald's work he says

> ... the fundamental frequency [of cries] initially rises. The fundamental frequency then remains relatively steady or gradually falls until the end of the cry, *when it typically falls at a faster rate.* [p. 41, italics mine]

This seems to be a misinterpretation of what Ostwald actually said, which was:

> After the peak [of the cries' F_o contour] there is a downward glide of pitch to the original pre-peak level. Some cries fall to a slightly lower pitch-level at the end. [p. 40]

By this Ostwald meant that most cries end at the same pitch level they began at, but with some the downward glide may end at a pitch level lower than the initial one (Ostwald 1969, personal communication). This may be represented graphically as follows:



Ostwald nowhere says that there is an abrupt fall or a faster rate of pitch fall at the very end of the cry.

Lieberman also cites the work of Wolff (1966) as providing support for the claim that attention-getting or "referential" cries

> all end with a falling fundamental frequency but that, in contrast, the noncry vocalizations often end with a level or a rising fundamental frequency contour. [p. 42, n.]

Wolff does note such a dichotomy but it seems that what he calls "noncry vocalizations" overlaps with what Lieberman calls "referential" or attention-getting cries. Specifically, Wolff notes that noncry vocalizations may occur when "the infant has no distress but simply 'wants attention'" (p. 98). However this is a small point. More significant is the apparent inter-subject variation that can occur with infant cries, because even if most of the infants in Wolff's study had a rapid drop in F_o at the very end of their cries, we have already seen from the study of Truby and Lind (1965) that not all infants do.

The foundation for Lieberman's conclusion that the F_o of infants' cries is a fortuitous by-product of their respiratory activity, is the following chain of assumptions:

> We shall try to avoid making any assumptions that do not seem warranted by the available physiologic and acoustic evidence. Where physiologic evidence is lacking, we shall assume that the infant uses the simplest pattern of articulatory activity when more than one pattern of articulatory activity could have been used to produce the same acoustic output.
>
> ... We shall assume that the infant does not precisely control the tension of his laryngeal muscles once phonation starts. He merely maintains the tension of the laryngeal muscles at or near the tension that they had as phonation started. This assumption is not particularly crucial, and the only pattern of laryngeal activity that we really must assume does *not* [italics his] take place is a consistent, controlled increase in the tension of the laryngeal muscles at the end of the breath-group. Our assumption is, however, consistent with the available physiologic data. [pp. 42-43]

The pledge with which this chain of assumptions begins is commendable. But in fact, none of the assumptions which follow is justified. The second assumption is certainly not warranted by the available physiologic data, and the first is surely by now a thoroughly discredited principle from which to argue in physiology or the behavioral sciences (this point is taken up below). It has already been argued above that a detailed examination of the evidence Lieberman cites provides no basis for concluding that the F_o contours of cries are mainly determined by their subglottal pressure patterns, even insofar as subglottal pressure may be considered a parameter independent of laryngeal adjustment, which, as pointed out above, is itself a highly doubtful assumption in dynamic phonatory conditions. Conversely, then, the only other possible determining factor of the F_o would be laryngeal activity. However, taking the laryngeal muscles to be involved in infant F_o control would have been a guess, too, if the only available work was the brief report in Bosma (1964) -- and their excellent pioneering study of infant cry seems to be the only source for comprehensive data on the several key physiological parameters accompanying cry production. At that time no safe conclusion was possible regarding the participation of the larynx and its associated structures in controlling the acoustic parameters of cries. But in their more extensive 1965 report, Bosma, Truby, and Lind provide striking evidence for the

involvement of the larynx in cries. The cine-x-rays they made of the 30 infants -- while not revealing much of the intricate laryngeal motions -- do permit observation of laryngeal movements parallel to the film plane, and, of course, the highly repeatable and individual acoustic shapes of the cries were also recorded. They conclude, in various places in their monograph:

> The actions of the larynx and pharynx essentially define the infant's cry, since the less discriminate trunk motions of respiration are more or less predictable from those upper respiratory actions ...
>
> Within the individual infant, these upper respiratory motions are repetitive in similarity, actually demonstrating a remarkable paucity of response patterns available to the infant under stress. [p. 64]
>
> The cry performance is a remarkably discriminate coordination; it also differs distinctively among infants so that each manifests in his cry (and thus acoustically) a motor identification of himself. [p. 90]

Thus on this evidence one would expect further data on the laryngeal activity associated with infant cries to show fairly consistent and repeatable motor patterns. One would be justified in suspecting that they were innately determined in some sense (but not in the sense which implies "human-universal"), but conclusive proof that the motor patterns were part of the infant's genetic make-up would presumably have to await comparison of the cries of a given group of infants with the cries of their offspring over several generations.

Further it is not possible to accept the minimal criterion proposed by Lieberman in the above-quoted passage for acceptance of his hypothesis. It is not enough that a consistent, controlled increase in laryngeal muscle tension never be found; rather one must find that the laryngeal muscles underwent no consistent change whatever that could affect the F_o at the ends of cries. *Consistent* must mean consistent within the cries of any single given infant. Thus to justify Lieberman's assumption it must not be the case that the laryngeal muscles maintain or even gradually vary the tension of the vocal cords throughout most of the cry and then suddenly release this tension at the end. Nor, of course, may it be the case that some muscle, such as the sternohyoid or another extrinsic laryngeal muscle attached to the sternum, consistently contracts at the end of the cry. Either of these actions -- although having the effect of lowering the pitch -- would falsify Lieberman's contention that the terminal F_o is determined by the subglottal pressure. In other words there

must be no consistent laryngeal action which affects pitch at the cry's termination. There is no direct evidence on this point yet, but the above remarks of Bosma et al. lead one to suspect that this criterion is not met.

Lieberman's assumption, that the simplest of two or more alternate ways of accomplishing an action is the one actually used, is, in principle unjustified. This may appear to be a standard procedure in science, but it is not. "Simplicity" is sometimes regarded as a well defined, useful, intrinsic quality of theories which enables one to pick one theory over another. This may or may not be so.* However this is not how Lieberman defines or uses the term, at least as far as can be determined by the above quoted context and elsewhere in his book. "Simplicity" in the above passage seems to mean something like "involves less effort" or "involves fewer muscles." This seems to be what he has in mind when he remarks that the tensing of the laryngeal muscles at the end of the marked breath-group is "in a sense the 'simplest' alternative to the unmarked breath-group" (p. 105). Lieberman appeals to the principle of simplicity not to decide between the simplest of two or more theories, but rather in an attempt to fill the gaps of knowledge. Whatever principle may win acceptance in physiology and the behavioral sciences in cases such as this, simplicity of muscular action would be immediately ruled out as a possible one.

For example, Lorenz offers evidence that inherited behavioral patterns often do not correspond to what might be termed "optimal" or "the simplest possible" behavior for particular species (Lorenz 1958). And one could argue that it would be simpler, in some sense, if human beings walked on all fours instead of just two legs: it would be more stable, the young of the species could acquire the skill sooner, etc. Similarly it would be simpler if those humans who can wiggle their ears didn't have their muscular repertory encumbered by this useless ability. Further, Person (1958) and Manukovskaia (1959) in electromyographic studies of the acquisition and refinement of simple manual and gymnastic skills have shown that any unskilled action is accomplished in a jerky, a-rhythmic, uneconomical way, typically involving many

---

* It probably is not so, as Popper (1959) has argued. There appears to be no way in which the notion of "simplicity" can be precisely defined and non-arbitrarily applied in any particular case where one of two alternate theories both empirically adequate has to be chosen on the basis of its intrinsic "structural" qualities. One may, as Popper does, equate the term with "degree of falsifiability," but we are led to prize this as a quality of theories for independent reasons and thus have no need for the superfluous and more easily misunderstood term "simplicity."

more muscles than are directly useful in the given task. Thus when confronted with an unfamiliar task, or when under stress, we frequently frown, clench our teeth and in general tense many more muscles than are needed to accomplish the task efficiently. A crying baby could quite appropriately be described as a throbbing bundle of tension. His whole body participates in the cry. Even if we were to assume that it would be simpler for the F_o of a cry to be determined by the subglottal pressure, we would still have no justification for assuming the crying infant would do things in the physiologically most efficient, i.e., in the simplest, manner.

Summing up, it would appear that we have as yet no evidence that one particular kind of infant cry is the progenitor of the declarative intonation or "unmarked breath-group" in speech. There is no evidence that one form of crying is simpler than another, thus entitling it to be labelled "unmarked." Thus any claims about the universality of certain aspects of intonation cannot be supported yet by evidence from infant cries.

Finally, if one looks to infants' vocalizations for a possible antecedent to speech, is it reasonable to regard cries as the most likely candidate? Are not infants' babblings or cooings -- in spite of their being less "referential" than cries -- more plausible as representing the beginnings of speech? Lenneberg (1967) specifically argues against the notion that the antecedents of speech must be "referential."

> There is evidence, although only of a circumstantial nature, that language does not emerge as a response to an experienced need, as a result of discovery of its practical utility, or as a product of purposive striving toward facilitated verbal communication. [p. 139]

In an earlier section Lenneberg presents the time schedule of language development in children and traces the gradual changes in *babbling* and *cooing*, not cries. The literature seems to have established that it is the intonation of the mother tongue which is acquired first by children -- for signalling emotion, etc. -- but, again, these intonation patterns are encountered in the spontaneous babbling of children, not in their cries. There is no reason to expect that the apparently invariant, reflexly-determined aspects of crying will be used in babbling, which typically manifests a great deal of creativity and elaboration of sound patterns. Also, given the wide variety of types of vocalizations that infants produce, we could isolate any one of them and thus find support for almost any hypothesis we might have about the origin of some aspect of speech. In fact anyone

proposing a hypothesis that certain elements of the neuromuscular programs employed in speech are innately determined must propose the crucial experiment which would decide between this hypothesis and the more traditional -- if, according to current tastes, less interesting -- hypothesis that these phonetic features of speech are *learned*, gradually, in imitation of what the child hears from those speaking around him.

Examination of the Evidence from the Hadding-Koch and Studdert-Kennedy Experiment

Lieberman cites the experimental results of Hadding-Koch and Studdert-Kennedy (1964) as evidence that speakers of Swedish and American English interpret intonation contours as if they manifested the acoustic correlates of the two posited features or phonological units [BG] (breath-group) and [P_s] (subglottal pressure peak), and further that they could take account of the physiological constraints of the vocal tract when decoding the intonation contours by using an "analysis-by-synthesis" scheme.

Briefly, the Hadding-Koch, Studdert-Kennedy experiment was the following. Forty-two complex intonation contours were superimposed on a carrier phrase "for Jane" (chosen to be acceptable to both English and Swedish listeners). The carrier phrase was vocoder processed and was identical for all the stimuli. The intonation contours were varied as follows: all began with a level pitch of 250 Hz, on the word "for" then rose on "Jane" to one of two possible high points (310 or 370 Hz); then fell to one of three possible "turning points" (130, 175, or 220 Hz), and finally went to one of seven possible "endpoints" (130, 145, 175, 220, 275, 310, or 370 Hz). The durations between these points of pitch change were identical for all the stimuli (see Figure 43 for samples of the test stimuli). The stimuli were played to 25 Swedish and 24 American students who judged them semantically and psychophysically in two separate sessions, categorizing them as statements or questions in the first, and judging whether they ended with falling or rising pitch in the second.

Without going into the detailed results of the two-category judgments, it turned out that those stimuli having the high points of 370 Hz required a smaller terminal rise to be heard as a question than did those with a 310 Hz high point. This effect can be illustrated by the graph in Figure 39 which plots the American's judgments (as o/o of "question" responses) against the terminal rise (in Hertz) for two series of stimuli differing only in their high points. Not all of the psychophysical results were published, but it appeared that when the lower of the two possible high points

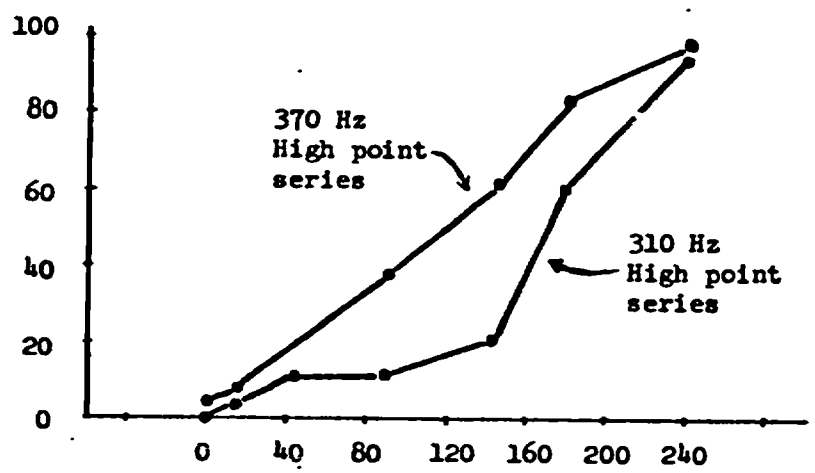


Figure 39. American listeners' responses to two series of synthetic intonation contours. Abscissa: endpoint minus turning point in Hertz (i.e., magnitude of terminal rise); Ordinate: percentage of "question" judgments. For a given terminal rise, the stimuli with the higher high point are judged more question-like than those with the lower high point. (Traced from Hadding-Koch and Studdert-Kennedy, 1964.)

were used (310 Hz) the subjects responded in this task much as they did in the semantic task, i.e., giving the response "question" for roughly the same stimuli to which they would respond "ends in rising pitch." When the 370 Hz high point was used, their responses were quite dissimilar in the two tasks.

Lieberman claims that the higher (370 Hz) pitch peak is interpreted by listeners as [+P_s], that is, the momentary increase in pitch due to the subglottal pressure peak, and that given the constraints of the respiratory system, this extra pressure early in the breath-group results in a lower subglottal pressure thereafter, vis-a-vis the same breath-group without [+P_s]. Assuming, as Lieberman does, that the subglottal pressure significantly affects the F_o in speech, this lowered subglottal pressure would cause the F_o after the peak to be lowered as well.

Then, assuming the contraction of the laryngeal muscles is the same at the end of all marked breath-groups, the resultant terminal pitch rise would then be lessened a certain amount. Listeners, Lieberman contends, nevertheless react identically to the different terminal rises (judging them both as questions), because, using an analysis-by-synthesis scheme they can "regenerate" the intended pitch rise even though it does not actually occur. This is because the listeners "know" (tacitly) what the constraints of the vocal apparatus are, i.e., that a pressure peak early in the breath-group depletes the air supply enough to lower subglottal pressure and consequently F_o in the remaining part.

Lieberman's analysis of Hadding-Koch and Studdert-Kennedy's data is neat, simple, and even intellectually satisfying. There is some precedent in current linguistics for accepting an analysis on these criteria alone. However, a close examination of the assumptions and data which Lieberman's analysis is based on will show that it is unfounded:

1. The existence of the hypothesized "Air Pressure Perturbation Effect" (henceforth, the APPE) is doubtful in view of what is already known about the dynamics of respiration in speech.

2. It is evident in Lieberman's data and that presented here that a lowering of the subglottal pressure after a pressure peak early in the breath-group does not invariably occur, and that when it does occur it cannot be attributed with any certainty to the cause Lieberman suggests.

3. His analysis logically rules out some intonation contours that do in fact exist.

4. Finally, there is another, better explanation of the results of the Hadding-Koch and Studdert-Kennedy experiment.

The APPE Finds No Support in the Literature

Lieberman's admittedly unverified hypothesis of why the "air pressure perturbation effect" occurs is the following:

> ... the physiologic basis of this effect may perhaps be a consequence of the fact that the elastic recoil force of the lungs is the main force that acts to expel air out of the lungs during each breath-group. The elastic recoil force is a function of the instantaneous volume of the lungs ...
>
> The presence of [$+P_s$] in the early part of the breath-group results in a greater airflow out of lungs, which, of course, lowers the volume of the lungs more than would have been the case if [$+P_s$] had not occurred. The elastic recoil force, ... therefore decreases. [p. 100]

Serious difficulties are encountered when verification for this hypothesis is sought in the literature. From the work of Rohrer (1916), Ladefoged (1962 and 1967a), Bouhuys, Proctor, and Mead (1966), Hoshiko and Blockcolsky (1967), and Mead, Bouhuys, and Proctor (1968), we know at least the following points:

1. The elastic recoil force of the lungs can be as high as 25-30 cm. aq., but only at maximum inspiration. In ordinary conversation after a normal inspiration the elastic recoil force is only about 5 cm. aq. or less. This would, theoretically, be sufficient to drive the vocal cords' oscillations for a short time -- perhaps a second or two -- until with less air in the lungs the recoil force dropped to below that required for phonation.

2. Ordinary speech -- whether spontaneous speech or the reading of a text in a consciously experimental situation -- typically involves the use of tidal air (the amount normally intaken) plus supplementary air (the reserve air in the lungs), so that many of the sentences uttered begin with very little positive elastic recoil and end with the elastic recoil force necessarily negative.

3. Although the elastic recoil force might be sufficient for phonation at the start of some expirations, in fact it is found electromyographically that most subjects use some of their expiratory muscles, the internal intercostals, from the very beginning of an utterance started on a normal lungful of air.

It would seem, then, that for many or even most of the sentences speakers utter the elastic recoil force of the lungs cannot be the main force in generating the air flow that drives the vocal cords.

The APPE Does Not Occur Consistently

Several pages after first proposing this APPE hypothesis, Lieberman qualifies it somewhat. He notes that this effect does not show up unless the pair of sentences, one with [+P_s] and one without, have roughly the same duration (within 100 msec.) and the same average subglottal pressure (within 1 cm. aq.). This may be reasonable, although one would have thought it almost impossible to ascertain whether any two sentences had the same average pressure in the case of Lieberman's data due to the admitted uncertainty about the accuracy of the records in the transient portions.* Further, it is difficult to imagine which point on a pressure curve containing a peak one would measure to see if it had the

---

* Lieberman states on page 62:

> For transient conditions the calculated subglottal air pressure must be regarded as an approximation of the true subglottal air pressure. When we made quantitative measurements that involved the subglottal air pressure, we therefore tried to use only the quasi-steady-state portions of the calculated pressure functions where the subglottal pressure was relatively steady.

In a footnote on the same page it is stated that

> The quasi-steady-state portions of the [P_s] function had variations of less than 0.2 cm. H_2O over 100 msec.

The non-"quasi-steady-state portions" amounted to at least half of each pressure curve.

same average pressure as another sentence without the prominence. If the prominence occurs at the beginning of the breath-group it is impossible to compare the pressures there, and, of course, one cannot compare the pressures after the pressure peak because those portions are supposed to exhibit the difference in subglottal pressure. It may well be that Lieberman himself followed a consistent procedure in these measurements, but the lack of details makes it difficult for anyone who wishes to replicate his findings.

It is also reasonable to look for the effect only when the durations of the sentences are comparable. If the effect is due to a more rapid than normal reduction in the lung volume as he suggests, the effects of extending the phonation should reduce the lung volume even more. This would enhance the effect if the sentence with [+P_s] were lengthened, but would obscure the effect if the sentence without [+P_s] was lengthened. Thus, the effect does not appear in his sentences 4.22 and 4.24 presumably because they are of different durations. But it is sentence 4.24 -- the one with [+P_s] -- that is lengthened, so there should be an enhancement of the air perturbation effect. There isn't, however. Nevertheless, out of all of his data Lieberman did manage to find 11 to 13 pairs of sentences which manifested the APPE and satisfied the stated restrictions.

In a different work Lieberman (1968a) presents the subglottal pressure curves for "Joe ate his soup" and "Joe ate his soup." Their durations seem to be equal so far as this can be ascertained from just the pressure curves (the zero point of the time scale is misplaced) and they have about the same average subglottal pressure. The APPE is not manifested; the pressure levels at the ends of the utterances are the same.

In our own measurements of subglottal pressure in the present study this effect failed to show up consistently. Table III presents some statistics on the pressure values at the end of phonation for ten tokens each of two pairs of sentence types. It will be noticed that some of the tokens in each pair differed in duration by slightly more than 100 msec. (155 msec. and 135 msec. in the worst cases). But no justification was given by Lieberman for the precise value of 100 msec., and here the extreme duration differences were in the direction favorable to an enhancement of the effect, i.e., the sentences with a [+P_s] were longer. In these records instances could be found of the pressure being lower at the end of the sentences with [+P_s], in accord with the APPE. However there are more cases where the pressure is equal or greater. The consistent appearance of this air pressure perturbation effect under the conditions specified has not yet been demonstrated.

Table III. Duration and Final Pressure for Two Pairs of Sentences Showing Lack of Air Pressure Perturbation Effect

	N	Pressure Mean at End of Phonation	Standard Deviation	Range	Total Duration Mean	Standard Deviation	Range
Joe ate his soup.	10	5.6 cm. aq.	0.4 cm. aq.	5.0-6.4 cm. aq.	990.0 msec.	30 msec.	940-1040 msec.
JOE ate his soup.	10	5.6	0.2	5.5-5.9	1040	35	980-1095
Bev bombed Bob.	10	4.6	0.3	4.1-5.2	1275	30	1215-1330
BEV bombed Bob.	10	4.7	0.3	4.1-5.0	1305	40	1215-1350

Studies of this sort actually do not mean much unless other variables are also controlled. Specifically, one must hold constant or at least keep track of: first, the action of the respiratory muscles (to see that they don't behave differently on the parts of the sentences to be compared) and second, the glottal resistance (which can be estimated if the subglottal pressure and the air flow are known). Both of these can affect the subglottal pressure independently of the elastic recoil of the lungs. While it is true that a subglottal pressure difference of the predicted sort does appear in some 13 pairs of sentences in Lieberman's data, and in some tokens of sentence-type pairs in the data presented above, it is also true that in a greater number of cases it did not show up or was in the wrong direction. Without other controls one cannot attribute the pressure discrepancy definitely to either larynx or lungs; nor if it is due to the lungs, can one say whether it results from fortuitous differences of lung volume and elastic recoil or from programmed muscular activity.

Logical Consequences of APPE Not Borne Out

If it were true that a momentary subglottal pressure peak produced lowered subglottal pressure and lower F_o thereafter, and if it were true that a pitch rise early in the breath-group is never due to an increase in laryngeal tension (Lieberman 1967a, p. 105), then prominent syllables early in a breath-group should always produce a lowered F_o contour after the prominence, as in Figure 40. (Since only the non-terminal position of a breath-group is affected, this should work for marked and unmarked breath-groups alike. Hence the terminal contour in Figure 40 is left unspecified.) But as has been indicated above there are clear cases in Lieberman's own data where this is not so, e.g., the sentences in his Figures 4.14 and 4.24. These are both questions and the contours quite clearly maintain a high level after the prominence on "Joe" as is shown in Figure 41. These patterns are difficult to account for in Lieberman's conceptual framework.

The APPE would also predict that any breath-group having more than one prominent syllable (assuming, of course, that each prominence was accompanied by a genuine "archetypal" subglottal pressure peak, and was not due merely to an "alternate articulatory maneuver" such as changes in vocal cord tension) would have a subglottal pressure curve and a F_o curve looking something like a cross-sectional view of the terraced rice paddies of Burma as in Figure 42. That this is not the case should be patently obvious to anyone familiar with English intonation. But to demonstrate it we need only look at the subglottal pressure curve for the phrase "Jóe ate a bíg bowl of bláck Borscht" (Figure 2, Lieberman

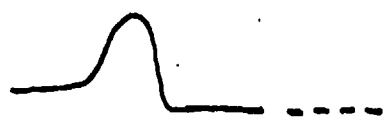


Figure 40.

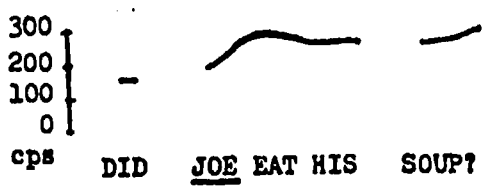


Figure 41a.

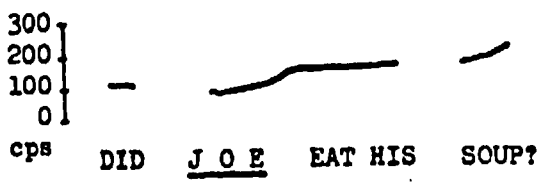


Figure 41b.

(Traced from Lieberman 1967a, Figures 4.14 and 4.24.)

Figure 42.

1968c). The pressure on "bowl" is higher than that on "ate" although the APPE should have produced the opposite.

Once it is admitted that laryngeal tension can and does control pitch changes in the non-terminal (as well as the terminal) positions of the American English breath-groups, then this particular objection to the APPE no longer holds; but this admission itself undermines the original basis for postulating the APPE.

There Is A Better Explanation of the Hadding-Koch and Studdert-Kennedy Experiment*

Apparently it was in order to explain the results of the Hadding-Koch and Studdert-Kennedy experiment (1964) that Lieberman posited the "air pressure perturbation effect" in the first place. But another interpretation of their experimental results is possible, one which in fact requires no reference to fortuitous variations in F_0 or sub-glottal pressure and one which need not posit any "analysis-by-synthesis" scheme on the part of the listener.

Before presenting this alternative explanation it is necessary to review, briefly, some of the facts of English intonation. In general this review is drawn from the insights of Palmer and Blandford (1964), Armstrong and Ward (1926), Pike (1945), Bolinger (1965) and Vanderslice (1968a and 1968b). These points have, to a large extent, been verified experimentally through speech synthesis by the last two authors as well as by Mattingly (1966, 1968). This is not to deny that there are differences in the treatment of intonation by these authors nor to suggest that any of them approaches completeness. There are also differences between British English and American English intonation, though not in the aspects relevant to the discussion. Terminology differs, of course, but an attempt will be made to make the meaning of the terms used clear by context or by example.

There are both lexical and sentence pitch modulations in English, that is, accent (or "stress"**) and the sentence intonational contour, respectively. If there is no prior context or discourse (which is the case when many amateur and some professional phoneticians conduct their introspective "experiments" regarding intonation and "stress") and if the speaker does not choose to emphasize any particular word (which would logically imply some prior context anyway) then it is the last accented syllable in the sentence which gets the major pitch contour

* This analysis owes much to some suggestions by Peter Ladefoged.

** See page 111, below, and Bolinger (1958) and Vanderslice (1968a) for the difference in these terms.

superimposed on (and/or after) it. Considering, for example, two possible contours, falling and rising, one can have the following sentences and accompanying F_o contours:

The old grey goose is dead. (1)

The old grey goose is dead? (2)

If there are any unaccented syllables after the contour, they will remain low pitched, and possibly tend to go lower still after the falling contour, whereas after a rising contour the tone will remain high and will go slightly higher at the end. Thus if the last word in the sentence had some unstressed syllables at the end, the intonation would be as follows:

The old grey goose is formidable. (3)

The old grey goose is formidable? (4)

If there is prior context, linguistic or otherwise, certain words, or rather the stressed syllables in them may be de-accented. A de-accented syllable will not get a pitch contour superimposed on it. For example, in response to the question, "When is Christmas?" it would be anomalous to answer

Tomorrow's Christmas. (5)

even on the 24th of December. The reply expected of a native speaker of Standard English, given the prior context provided by the question, would be

Tomorrow's Christmas. (6)

Specifically, "[i]s Christmas" is old information and is accordingly obligatorily de-accented. Similarly, the context set by such sentences as "Which grey goose is dead?" (7) or "Here lies the young

grey goose that has gone to its final reward" (8) would require that sentence (1) be rendered

The old grey goose is dead. (9)

Likewise sentence (2) as an echo question to (and therefore in the context of) sentence (9) or (1) would have to be rendered

The old grey goose is dead? (10)

Whether the syllables following the pitch contour are lexically unaccentable or unaccented due to accent deletion it is still the case that the tone on them remains low after a falling contour and remains high tending a bit higher still, after a rising contour.*

Palmer and Blandford (1924) indicate that this small pitch rise that follows a non-terminal pitch contour is something that always occurs (p. 15). Pike (1945), on the other hand seems to claim that it frequently, but not necessarily always, occurs (p. 73-75).

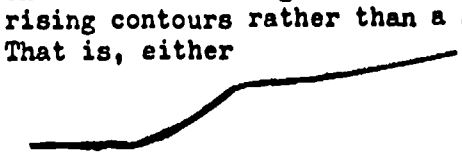
If the speaker chooses to emphasize or make prominent an accented syllable one thing he may do is superimpose a more noticeable pitch modulation on the emphasized word. Other contours are possible but let us consider just two, a large pitch obtrusion up in sentences with falling intonational pitch pattern and a large pitch obtrusion down in sentences with rising intonation. With emphasis, sentences (9) and (10) could be rendered

The old grey goose is dead. (11)

The old grey goose is dead? (12)

---

* The pitch after the rising contour may be made up of several little rising contours rather than a single long rising contour. That is, either



or

This traditional analysis of intonation contours would lead us to expect that a given contour would be more likely to be identified as a "question" if (1) it had a large rising pitch at the end, or if (2) it had a large rising pitch *before* the end and (a) thereafter remained high and possibly (b) had a slight rise at the end. This is exactly what the results of the Hadding-Koch and Studdert-Kennedy data reveal. When the pitch rises to the 370 Hz high point this is indeed interpreted by the listeners as "prominence" on that syllable -- Lieberman is right on this point. However, what Lieberman does not recognize is that a pitch rise early in the utterance due to prominence or emphasis manifests the major intonation contour of the sentence. This accounts for the fact that a smaller terminal rise is sufficient for the 370 Hz high point stimuli to be judged as questions. This first large pitch rise is in itself a strong cue for "question" and seems to tip the perceptual scales in favor of that judgment; another large pitch rise at the end is not necessary for the contour to remain "question"-like. In order for a stimulus with such a large pitch rise to get a "statement" judgment it is necessary that there be an extra low pitch fall immediately thereafter and that the pitch remain relatively low.

As Hadding-Koch and Studdert-Kennedy noted (p. 180), for any given high point value, the higher the "turning point" the more will the stimulus appear to maintain a higher pitch level after the early rise and thus the more "question-like" it will be judged. This effect can most dramatically be illustrated by the two contours in Figure 43, both of which are identical in high point (370 Hz) and terminal rise (90 Hz) but differ in turning point. The one on the left, having a 220 Hz turning point was judged a question 85 °/o of the time, whereas the other having a 130 Hz turning point received only a 39 °/o score (among the U. S. listeners; the differences were even more dramatic with the Swedish listener's responses). This same turning point effect also shows up in the results of another recent experiment involving listeners' responses to synthetic intonation contours (Majewski and Blasdell 1969). This effect is accounted for by considering that after a single early rise (prominence) the pitch in a question typically remains high until the end; the synthetic contour with a high turning point more nearly approximates this situation. It would appear however that Lieberman's analysis does not and can not account for this turning point effect.

As has been suggested by Hadding-Koch and Studdert-Kennedy (1964, p. 184), this analysis may explain some of the cases of disagreement between the semantic and psychophysical responses to the same stimuli. The task of the listeners in the psychophysical test was to tell if the final pitch contour was rising

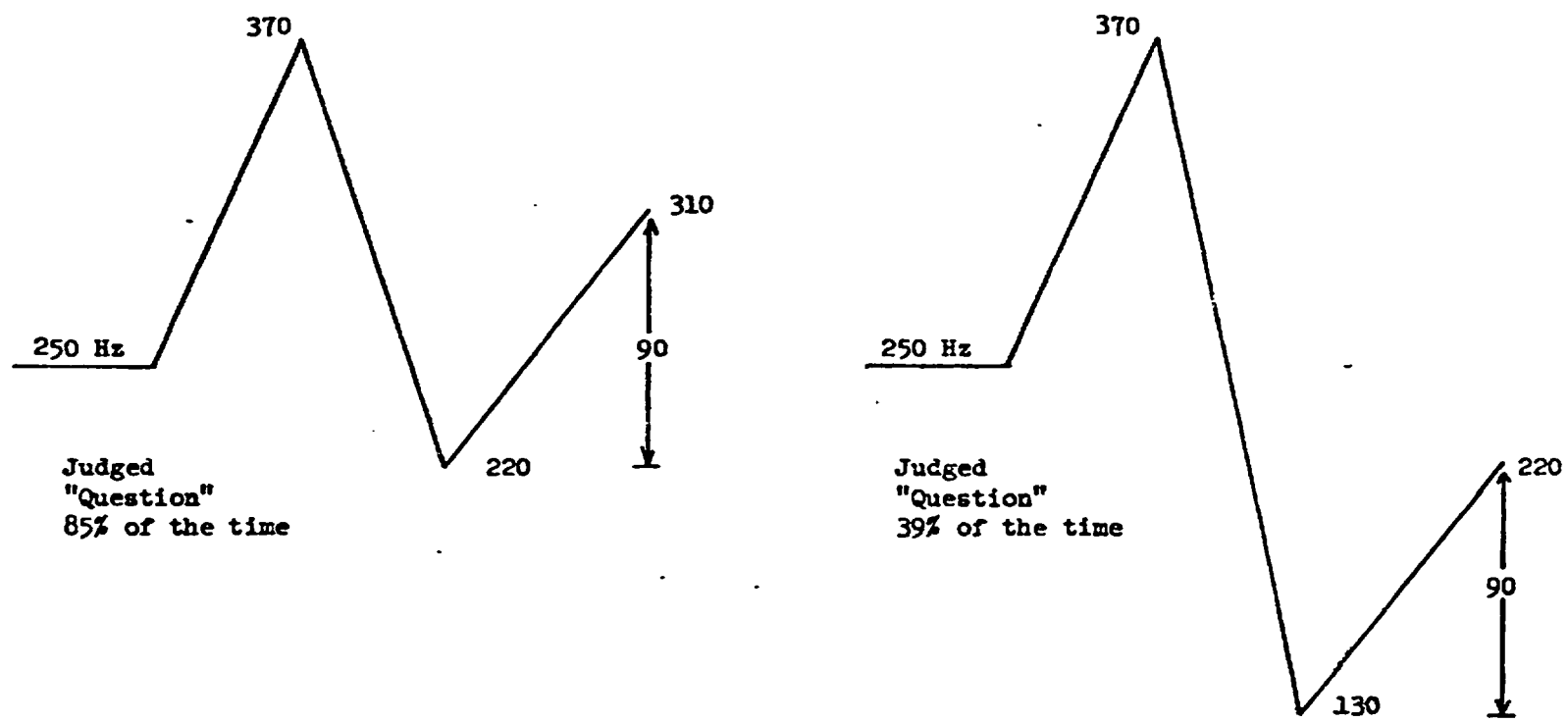


Figure 43. Two intonation contours having identical high points (370 Hz) and identical terminal rises (90 Hz) but which received different percentages of question judgments due to differences in the turning point. (After Hadding-Koch and Studdert-Kennedy, 1964)

or falling, but the task of judging stimuli as "questions" or "statements" may, if there is a large early pitch rise, have little to do with the *final* pitch rise. Thus, the two tasks may sometimes be entirely different for the same stimuli and we would expect on those occasions that their responses likewise would deviate. Only a fraction of the data from the psychophysical tests was published in the authors' 1964 article, but those records do seem to support this view. This is not to deny that there are undoubtedly many other "effects "to be discovered in this data explaining why the responses went this way or that way. Future experiments or closer scrutiny of the data may reveal these other factors which influence the identification of intonation contours as "question" or "statement."

These remarks on the results of this experiment are scarcely innovative; they are essentially anticipated by those of Hadding-Koch and Studdert-Kennedy themselves. The point here is simply that those experimenters' observations on their own quite excellent study are quite in accord with the traditional analyses of English intonation, (e.g., those of Palmer and Blandford, and of Pike), and require no reference to any "air pressure perturbation effect" or any "analysis-by-synthesis" on the part of the listeners.

From the preceding discussion it would appear that evidence has not yet been presented that the posited features [±BG] or [+P_g] have the kind of perceptual reality attributed to them by Lieberman nor is there any evidence yet that the feature [-BG] is innate.

Then how *can* we explain the universality of the falling and rising pitch contours? First, the universal *presence* of these contours in the languages of the world needs no complicated explanation. Pitch is a scalar phenomenon, it can either rise, fall or remain level; it is not at all surprising that all the possibilities for varying pitch actually do occur. What is surprising is the apparent uniformity of usage or "meaning" of the various contours in the world's languages, specifically, that rises or sustentions of pitch signal a questioning or uncertain attitutde on the part of the speaker, whereas a falling pitch indicates confidence, certitude, etc. Bolinger (1964) and others have pondered this point but have not, I believe, offered very convincing explanations. For this reason, Lieberman's provocative and highly interesting theory that certain aspects of intonation are innate would be very much valued if it was supported by any convincing evidence. But it is clearly unfounded. We could just "explain away" the problem by suggesting that when speech began humans made an arbitrary decision that pitch rises would mean one thing and pitch falls another and that we've

been stuck with that decision ever since. Or we could go the evolutionary route and note that animals (with larynges) usually make high-pitched whining noises when they are afraid or uncertain of their safety, whereas they produce low-pitch growls when they are angry or aggressive, and from this we could extrapolate to intonation in speech and claim that humans use roughly the same code. If some such reasons as these do underly the uniformity of usage of pitch contours in the world, it is unfortunate because they will be very difficult to prove. It is more attractive to take the approach Lieberman did and seek a physiological explanation for the problem. But unfortunately the existing evidence just does not fit and science doesn't sanction the acceptance of a theory merely on the basis of its attractiveness.

* * * *

On the Perception of "Stress"

There have been some recent provocative claims made by generative phonologists concerning the perception of "stress" and its role in intonation (Chomsky 1967b, Lieberman 1967a, Chomsky and Halle 1968). Although this is only marginally relevant to the preceding investigation concerning the control of pitch in speech, these claims deserve some examination since the requirements of the listener should be taken into consideration in developing a model of any aspect of speech production.

The fact which the generativists' model of stress perception was designed to explain is the putative ability of phoneticians to categorize the syllables in novel utterances in terms of multiple levels of stress -- of the Trager and Smith (1951) variety -- for which levels of stress there is or need be no physical basis in the acoustic speech signal. In Lieberman (1967a) and Chomsky (1967b) it seems that at least two levels of stress are allowed to be physically present; in Chomsky and Halle (1968) rules are given for predicting all levels of stress and the impression is given that even if there is a physical basis for differentiating some stress levels, such acoustic cues are not really necessary. The ability of phoneticians to regenerate "missing" levels of stress is explained by their knowing, tacitly, the stress assignment rules of English (based on the principle of the transformational cycle) and their identification of the lexical items

and their syntactic relations in the heard utterance.* Lieberman 1967a refers to this process as an "analysis-by-synthesis" technique (p. 162), but more generally it is simply called "internal computation" or a similar term.

This claim about the perception of stress is an important point in linguistic theory because on it rests the principle justification for the transformational cycle in phonology, which in turn (together with the supposed application of the transformational cycle in syntax) serves as the prime example of what must be part of the "innate language acquisition mechanism" (it being alleged that the transformational cycle is too complicated to learn in the way that ordinary language-specific features are learned). And all of this finally is offered as demonstrating that there is substance to the phrase "linguistic contributions to the study of mind." For conclusions as important and as far-reaching as these are, one would expect that the low-level facts basic to the theory would be rigorously substantiated. Chomsky and Halle, however, believe the "facts" about stress perception are so well known and so firmly and unquestionably established that they need give no attention to substantiating them. Thus they repeatedly assert that "careful phoneticians trained in the same conventions" can easily categorize the syllables in novel utterances in terms of several degrees (from 4 to 6 or more) of "stress" and state that this "has long been known" (p. 59) or that there is "no doubt" (p. 24) or "there is ample evidence" (p. 25) that this is the case, and other such remarks. This, in spite of the admission that

> ... it is not surprising that there should be great difficulty, within impressionistic phonetics, in determining how many stress levels should be marked and how they are

---

* This is the only interesting formulation of the claim, that is, that the "trained linguist" need only have a *tacit* awareness of such rules; that his training has merely enabled him to "bring to the surface" his subconscious knowledge of these stress rules. An alternate but wholly uninteresting formulation of the claim would be that only those linguists who formalized the stress rules and who use them in the classroom every day were capable of identifying the multiple levels of stress with ease. It has never been made entirely clear by the generative phonologists exactly to which subset of the total population of linguists this claim is supposed to apply. This is not a trivial matter: with appropriate disregard for choice of experimental population we could quite easily prove the psychological reality of partial differential equations by using as our test subjects experts on partial differential equations. And since such complex mathematics are even more difficult to learn than the transformational cycle we could easily demonstrate that it would have to be innate, too.

> distributed in utterances that exceed a certain degree of complexity [p. 25]

and in spite of testimony and evidence to the contrary by many other noted linguists.

There is, for example, Sledd's (1955) comment in a review of Trager and Smith's *Outline of English Structure*:

> Anyone who has attempted to analyze or teach the English patterns of pitch and stress knows that competent observers may vigorously disagree and a single observer may disagree with himself so often as to make secure confidence in his own judgements painfully difficult; this reviewer, at least, simply does not hear the neatly symmetrical distribution of pitch allophones with phonemes of stress as Trager and Smith describe it [and] ... he is sometimes in serious doubt whether to write primary or secondary stress ...

Furthermore, the First and Second Texas Conferences (Hill 1962a and 1962b) contain numerous testimonies and anecdotes revealing that considerable disagreement exists between noted linguists in their stress assignments to English phrases.

Lieberman (1967a), although fully supporting the notion that linguists can differentiate multiple levels of stress, nevertheless allows that

> The results of a number of independent psychoacoustic experiments suggest that listeners can make only binary categorical distinctions along the dimension of prominence when they listen to connected discourse.

and then adds in a footnote:

> ... Hadding-Koch (1961) found that the stress levels assigned by listeners, other than level 1, were randomly distributed. ... Lisker, in a personal communication, notes that students of the Trager-Smith notation consistently produce random results with respect to differentiating the intermediate stress levels. [p. 149]

Also there is the published discussion during the conference on "Linguistics and the Teaching of English as a Foreign Language," held at Ann Arbor, Michigan, in the summer of 1957 (*Language Learning* 1958), in which it appears that upon a request by Albert Marckwardt for a show of hands of all those linguists attending who had difficulty distinguishing even four levels of stress, about half of those present raised their hands. This event was noted and commented on by Wang (1962)

whose comments were in turn noted by Postal (1968). Postal in the course of flogging the "autonomous phonemicists" elaborates on this point and notes that

> ... the evident inability of English speakers to recognize and discriminate the four stress phonemes ... has been admitted even by the most ardent supporters of the four-stress system ... [p. 25]

He goes on to testify that

> ... it is my experience that for most speakers, as with myself, any distinction beyond those of primary versus nonprimary are difficult to hear (even with some phonetic training), and even this distinction is not without its difficulties. [p. 26]

Postal still maintains, though, that

> ... contemporary systematic phonemics [i.e., generative phonology] requires representations of English in which no stress distinctions are marked and hence correctly predicts that untutored speakers [which by his own testimony includes Postal himself] should have the actually observed difficulties with stress distinctions. [p. 27]

These views being quite well known in the field, would indicate at the very least that it is far from being so firmly and unquestionably established that trained linguists or anyone else can unerringly and easily distinguish multiple levels of stress on the syllables of English utterances.

Chomsky and Halle rightly acknowledge that these issues are empirical matters and as such are subject to experimental verification (p. 26) but they unfortunately undermine in advance the significance of the results of any test should anyone ever bother to devise and run one. They say that since these stress contours are not physically represented in the acoustic signal,

> There may be no empirical sense to the question of whether the resulting representation is correct in full detail [p. 25]

> ... Furthermore, the representation of the perceptual facts is likely to be governed in part by arbitrary convention irrelevant cognitive limitations after a certain degree of complexity is reached. Thus, it is impossible to expect (and, for purposes of investigating linguistic structures, unnecessary to attain) a complete correspondence between the records of the impressionistic phonetician and what is predicted by a systematic theory that seeks to account for the perceptual facts that underlie these records. [p. 26]

The purveyors of the new imperial wardrobe (Andersen 1837) could not have put the argument better. In other words, there is no way either of knowing or testing the reality of these posited stress contours; we must simply accept the word of those who posited them that they are accurate and true. The above-quoted escape clauses provide the theory with perfect insulation against any possible disconfirmation. They can never be justified nor accepted if linguistics is sincere in claiming to be an empirical discipline.

There is, in fact, some non-anecdotal evidence marginal to the issue of whether or not linguists can consistently recognize several levels of stress on the syllables of English utterances.

The experiment reported by Lieberman (1965) showed that two trained linguists were not consistent in marking syllables of utterances with the four *pitch* levels à la Trager and Smith. However only one of the two subjects marked the material for "stress" so there is no chance to determine the consistency of responses between linguists in this task. Also in both tasks the linguists marked the material differently when they heard it in an acoustically distorted form (retaining only pitch and/or intensity information) than when they heard it in full, undistorted form. Chomsky and Halle state that this result "strongly suggests that what the phonetician 'hears' in utterances depends very heavily on internalized rules that predict perceived phonetic shape" (p. 26, fn.). This hardly exhausts all the possibilities. There are at least the following reasons why linguists might make different judgments of pitch and stress contours when listening to natural speech as opposed to acoustically distorted speech:

a. They use information on the lexical and syntactic relations in the sentence plus internalized rules such as those in *Sound Pattern of English*. This would be an "interesting" version of Chomsky and Halle's interpretation.

b. In making judgments on the natural speech they use syntactic information to "type" the sentence according to a given taxonomy, say Trager and Smith's, and then assign the pitch and stress contours decreed by the taxonomic system for those sentence types. This is the interpretation Lieberman offers for those few cases in which the two linguists were consistent in the judgments of pitch levels. This would involve "internalized rules that predict phonetic shape" but would be a rather uninteresting version of such.

c. They react to the same cues in both tasks, that is, amplitude and/or pitch variations, but having had a lifetime of practice in judging these parameters in speech situations, they come with a perceptual bias when judging them as speech but do not have this bias when judging them as non-speech (Lane 1965 and 1967, Broadbent 1967).

d. In making their judgments they use the other phonetic cues present in natural speech but not present in the distorted speech signal, e.g., vowel quality, degree of aspiration, durational cues, etc.

e. They are subject to phonetic fantasies which are completely idiosyncratic and unstructured. That is, we might expect much the same kind of results if the subjects took the test while under the influence of LSD.

f. Some combination of the above reasons -- or other reasons.

We might be able to accept the first interpretation if the two native speaker linguists agreed with each other in their markings, however it was clearly shown by Lieberman that they were not consistent. Unfortunately the experiment lacked sufficient controls to allow us to pick out any of the other alternatives.

Ladefoged and Fromkin (1968) report an experiment administered to 24 subjects attending a linguistics lecture, which presumably would imply that they were linguists although their degree of training was not indicated. The subjects were asked to indicate via a phonetic transcription their pronunciation of 12 orthographically-presented English-like nonsense words, each used in a possible English sentence, e.g., "He is going to sitrenide the paper" and "Then he will sempit the result." The results, reported for 11 of the 12 test words, indicate a reasonably good, but not perfect, agreement between the subjects in recognizing up to 3 degrees of stress, although unfortunately in giving the results the authors did not count "minor differences in stress such as that between levels 2 and 3 in the Trager and Smith system." (In Lieberman's experiment such differences in the markings of the pitch levels were counted and they contributed to the inconsistency between the two subjects.) In spite of this and the fact that strictly speaking this was not a perceptual test, it does confirm that linguists -- or speakers of English in general -- are fairly consistent in providing a phonetic shape to orthographically-presented words, using, presumably, some procedures which could also be available for speech perception. Of course this offers no evidence that such procedures are actually used in speech perception and for the reasons put forth in Chapter One, it offers as yet no evidence that the procedures used are best described by the stress assignment rules given by Chomsky and Halle.

To sum up the situation so far, it appears that it has not been established convincingly that trained linguists or anyone else can easily, accurately, and consistently recognize multiple levels of stress on novel English utterances by using an analysis-by-synthesis scheme or a kind of internal computation using rules incorporating the transformational cycle. Thus all of the elaborate theorizing based on this point is without any firm foundation.

But what is stress? If we are to run the crucial experiment just how are we to instruct our subjects? What shall we tell them to listen for in the utterances? Chomsky and Halle seem to assume that stress doesn't or needn't have any acoustic or physiological correlates. And

so in *Sound Pattern of English* they do not bother to mention what the correlates of stress might be. If all or some levels of stress are completely perceptual entities, that is, not intended as a part of the transmitted message by the speaker and thus not present in the acoustic signal, but are instead always produced entirely in the listener's brain, then this is a reasonable way to proceed. But this leads logically to a very curious conclusion. If the listener has to first identify the syntactic and lexical information in the sentence before he can generate the missing stresses, and if we agree that the listener has "got the message" or is well on his way to getting the message by the time he knows the words in the sentence and their syntactic relations, then the stresses contribute absolutely no information to the message. They would always be completely superfluous in a way that other so-called redundant parts of the message are not: e.g., the number of the grammatical subject is redundantly reflected in the verb in English in some tenses, but if a listener misses one of the cues for the number of the subject, then the remaining cue gives the necessary information. These stress levels, being unessential for the transmitted message can never provide such redundant cues. One may wonder why any listener -- trained linguist or not -- would bother to crank out these completely functionless and imaginary "phonetic details"?

This rather fantastic conclusion can be avoided if attention is paid to the existing literature. It is difficult to understand how anyone could write in the 1960's on the problem of stress perception and almost completely ignore the vast amount of experimental work that has been done on the issue. There is in fact a sizeable body of literature that discusses the acoustic and physiological correlates of what linguists have labeled "stress" and the functional role of this entity. Chomsky and Halle's sole references on this are the above-quoted interpretation of Lieberman's (1965) experiment and the two related statements:

> ...there is no evidence from experimental phonetics to suggest that these [stress] contours are actually present as physical properties of utterances in anything like the detail with which they are perceived. [p. 25]

and the footnote:

> ...however, even if such differentiations did exist along a single dimension of the acoustic signal, there would be some reason to doubt that they might be identified by phoneticians. There is evidence that even under experimental conditions, where complex stimuli are to be sorted along several dimensions, more than two or three distinctions along each dimension will overload the perceptual capacity ... [p. 26]

Why do the correlates of stress have to be limited to a single acoustic parameter? Anyways, being partial truths, these statements are distortions

of what experimental phonetics has actually shown about the thing linguists call stress, namely, (a) that linguists' first guess as to the physical correlate of "stress" was wrong, with the result that since they had been looking in the wrong place, naturally they could find no physical evidence for it, and (b) the real physical correlates of stress are multiple and complex, which helps to explain some linguists' *impression* that they could recognize multiple levels of stress: they were undoubtedly lumping together variations in more than one parameter. Subsequent linguistic analysis of the phenomena showed that (c) far from being functionless, manifestations of stress serve a very important role in speech.

It will be useful to attempt a brief summary of these points; for more systematic and detailed accounts of what "stress" is and how it functions, see Bolinger 1958 and 1965 and Vanderslice 1968a and 1968b.

Linguists originally thought that the difference between the noun and verb forms of "insult" was due to a difference in the placement of an increased breath force on the syllables of the word. They called this increased breath force (or articulatory force) "stress" and imagined that the acoustic correlate of this difference was increased loudness (Sweet 1911, p. 465). Even so, most linguists, as noted by Lieberman, recognized only two levels of stress, namely stressed and unstressed. Some linguists, for various reasons, convinced themselves and their students that they could recognize 4 levels of stress. Trager and Smith (1951) are notable for this and Chomsky and Halle follow their lead in this matter without question, although go even further by recognizing at least 6 levels of stress. Also, Chomsky, Halle, and Lukoff, along with Trager and Smith, accepted the common definition of stress as "degrees of loudness." However in *Sound Pattern of English* Chomsky and Halle carefully avoid suggesting that there may be any acoustic correlates of stress.

It's an old story now -- in fact, it was something of an old story in 1951 when *Outline of English Structure* appeared -- that the phoneticians gradually realized their concept of stress was wrong. Early investigations by Coleman (1911), Muyskens (1931), Scott (1939) and others were sufficient to make many noted linguists accept the idea that stress differences were primarily a matter of pitch change. More recent work by Lawrence (1952), Bolinger (1958), Fry (1955 and 1960), Lieberman (1960), Mattingly (1966), Rabiner, Levitt, and Rosenberg (1969) and many others has confirmed that what linguists label "stress" differences are largely due to variations in pitch, duration, vowel quality, and intensity, with pitch usually being the most effective cue. This by itself helps to explain why it was not possible to find a single parameter in the speech signal, such as intensity, which varied in a way paralleling the linguists' judgments on stress levels.

Bolinger further showed that it is not pitch levels which correlate with "stress" but simply pitch contours or obtrusions, either up or down -- which he called "pitch accent." Following Palmer and Blandford (1924) and others, Bolinger pointed out that a distinction must be made between the *potential* of a syllable to bear this pitch accent and the actual manifestation of the pitch accent. The potential to receive pitch accent

he designated "stress." In order to avoid terminological confusion "stress" meaning "potential for receiving pitch accent" will be given hereafter as "$stress_b$," and "stress" meaning "perceived relative prominence or loudness" will be given as "$stress_{ch}$."

Among the pitch contours that can serve to manifest a $stressed_b$ syllable is, of course, the main intonation contour (or "tune") that must be a part of every sentence. As pointed out above (pp.98-100), this sentence intonation contour is superimposed on the last syllable capable of receiving it. Some $stressed_b$ syllables' potential to bear pitch accent may not be realized due to contextual constraints. For example, the second syllable of "incarcerate" is the $stressed_b$ syllable and ordinarily its potential for pitch accent is realized. This is the case when it is pronounced in isolation, where, of course, it is the sentence intonation contour that is superimposed on the second syllable:

incarcerate.

However, in some contexts this potential may not be realized, specifically when the word represents old information, that is, is part of the "topic" as opposed to the "comment."

Instead of jailing the thief, it's the mayor we should incarcerate.

Here "jailing" and "incarcerate" are synonymous so "incarcerate" is old information and is accordingly "de-accented," that is, its potential for receiving pitch accent is not realized. "Mayor" is thus the last accentable word and so its first syllable receives the sentence intonation contour. The word "incarcerate" now falls in the level low-pitched portion (called "cadence" by Vanderslice) that necessarily follows this type of intonation contour. No pitch contour falls on its $stressed_b$ syllable.

An understanding of the mechanics of such aspects of intonation as "cadence" and placement of sentence intonation contours may shed some light on the origin of linguists' impressions regarding what they have called "$stress_{ch}$." Consider, for example, the following arm-chair experiment from Chomsky, Halle and Lukoff (repeated in *Sound Pattern of English*, p. 116, but without any mention of "loudness"):

> Consideration of single words reveals at least five phonetically distinct levels of stress. E.g., the relative stresses
> 3 4 1 5
> in "emendation" must be marked "emendation." We arrive at this conclusion by noting that the heaviest stress is on the third syllable, and the first syllable is clearly louder than the second, which in turn is clearly louder than the fourth.

> Furthermore, we note that the stress on the first syllable is less than that on the first syllable of "either nation", the stressed syllables of which must be marked 21; so that the first syllable of "emendation" must have stress 3.
>
> An even larger number of stress distinctions can be discovered when we proceed to longer phrases. [p. 70]

Whether the third syllable of "emendation" need be the loudest syllable is debatable, but it is unquestionably the stressed$_b$ syllable (any dictionary reveals this) and consequently is eligible to receive pitch accent. In this case since it is uttered in isolation it gets the sentence intonation contour. Common renditions might be:

emendation

or

emendation

The fourth syllable, "-tion," necessarily is pronounced during the low-pitched cadence. Since it is lower in pitch than the first two syllables it is natural that it seem relatively less prominent. However, we could simply attribute the same degree of stress$_b$, namely "not stressed " to the first, second, and fourth syllables and still *explain* quite naturally and simply the *perceived* greater prominence of the first and second with respect to the fourth.

There is indeed a quite compelling impression that the third syllable is more prominent than the others. This could sometimes be due to the higher pitch it may get. However syllables need not be uttered on a higher pitch than the others in order to be heard as the most prominent syllable because, as Vanderslice has pointed out, it is sufficient that the syllable receive the sentence intonation contour in order to stand out from the other syllables. This also explains why the second stressed$_b$ syllable of "either nation" appears to be more prominent than the first. In a common rendition, e.g.,

either nation

the stressed$_b$ syllable in "nation" is the last in the utterance and so receives the sentence intonation, thus appearing more prominent than the stressed$_b$ syllable in "either." But there would seem to be no need to formally attribute different degrees of "something" to either syllable;

the perceptual difference follows naturally as a by-product of the mechanism of the intonation system and the listener's hearing mechanism. Analogously, the visual impression received when viewing railroad tracks that the two rails meet at the horizon (when the line of sight is about parallel to the tracks) does not require that the tracks be constructed such that the rails actually do touch in the distance. Nor does it require any special innate cerebral ability on the part of the viewer. The impression follows naturally as a by-product of the way the world is constructed and the way vision operates. Thus, in speech neither the speaker nor the listener has to "know" about stress_{ch} in order for the impressions of relative prominence of syllables to occur. If this is the case, Chomsky and Halle's Nuclear Stress Rule (p. 90) which would give the stress contour 21 for the phrase "either nation" would thus be unnecessary or at least would say nothing about the mental processes of speakers of English.

No doubt the relative durations, the degree of vowel-consonant coarticulation, and other such matters also play a part in determining the perceptual impression of the relative prominence of the syllables in utterances. Actually without suitably controlled experiments it is impossible to know with any certainty why some linguists have these impressions or whether they all have the same impressions. Therefore the above comments should be regarded as hypotheses that need further testing. Nevertheless it is better to look for physical criteria the linguist might use to identify stress levels rather than to assume he must be "hearing things."

The algorithms or procedures the native speaker *does* need to know -- those which will fortuitously give the impressions regarding stress_{ch} -- are the ones which will correctly produce the pitch variations, the durational characteristics, etc. of utterances. Such rules can be and have been written with varying degrees of success (Kelly and Gerstman 1961, Holmes, Mattingly, and Shearme 1964, Kim 1965, Mattingly 1966, Vanderslice 1968a, Matsui, Suzuki, Umeda, and Omura 1968, Rabiner et al. 1969) and when used to synthesize speech produce quite satisfactory results -- results, which while not yet perfect, are impressive enough to indicate that they are on the right track.

Distinct from rules or procedures for generating the actual phonetic parameters of utterances, but also needed by the native speaker, are means whereby he knows which syllables are "stressed_b," that is, have a potential for pitch accent, which, again, is not the same thing as perceived prominence. From the preceding discussion it appears that stress_b-generating algorithms, if feasible, need work only on individual lexical items, not phrases and not sentences, and further need predict only two levels of stress_b not five or six. Those of Chomsky and Halle's stress rules that apply within the word could apparently accomplish this for a large number of English words, with appropriate trimming of the extra stress_{ch} levels. Of course the task could also be done by a kind of dictionary look-up (cf. Vanderslice 1968a). Since the look-up need be

done only for lexical items it presents no great burden on memory any more than the other phonetic characteristics of words do. However, whether the speaker knows that the third syllable of "emendation" is the stressed syllable by means of generative stress rules or by a kind of dictionary look-up is entirely an empirical matter that has yet to be settled.

In the light of the foregoing points let us consider one final bit of evidence alleged to show that listeners generate internally their own impressions of relative prominence irregardless of what is physically present in the acoustic signal. This is an experiment reported by Lieberman (1967a, p. 155). He began with the assumption that the two noun phrases "lighthouse keeper" (someone who tends a lighthouse) and "light housekeeper" (a small person who keeps house) are innately endowed with different stress levels, as shown. He then recorded someone saying the following two sentences which incorporated those noun phrases:

1 3 2 4
2 1 3 4

1. The life of a lighthouse keeper formerly was very lonely.

2. Our maid weighed 180 pounds, but the Joneses had a light housekeeper for more than twenty years.

He excised the two noun phrases from the sentences and interchanged them, i.e., putting the part excised from sentence 2 into sentence 1 and vice versa. The result was:

> It is ... impossible to hear any difference between the stress patterns of the altered sentences and the original sentences. In sentence 1, one hears the phrase as (*light house*) (*keeper*) whether or not the original sentence or the sentence with the phrase excised from sentence [2] is heard. The context of the entire sentence indicates the appropriate constituent structure, and the listener "hears" the correct stress pattern.

Lieberman fails to note that this result is due to the fact that in sentence 2 there is a synonymy relationship between "maid" and "housekeeper" which requires the de-accentuation of "housekeeper" and that there is a contrast between "weighed 180 pounds" and "light" which requires that the accentuation on "light" be manifested and possibly emphasized, thus giving the pitch pattern

light housekeeper.

This is approximately the same pitch pattern for "lighthouse keeper" in sentence 1. And so the two noun phrases can be interchanged in the two

sentences without noticeable effect. Thus the rules of accentuation as discussed above help to account for the homophony of the two phrases. Using the same principles of accentuation it is a simple matter to construct an alternate sentence to (2) which will give a very different pitch pattern to the noun phrase "light housekeeper." Instead of sentence 2 one need only ask that the following sentence be used and the same tape splicing procedure be used:

> 2'. We have an underweight gardener whereas the Joneses have a light housekeeper these days.

In this case "underweight" and "light" are now synonymous and this forces "light" to be de-accented, but "gardener" and "housekeeper" contrast requiring the (emphatic) accentuation of "housekeeper." The resultant pitch pattern will be

and this will, of course, sound quite anomalous in sentence 1, or at least will change its meaning. Neither the context of the entire sentence nor any amount of internalized rules will make the substitution sound like the original sentence.

The main point here, which is that made by Vanderslice, is that such items, the two forms of "light-house-keeper," "black-bird," "red-coat," "green-house," "grand-father-figure," etc., can all be rendered homophonous or non-homophonous quite naturally by providing the appropriate prior context. An interesting example from Vanderslice (1968b) is the following:

> Castor: The actor we want has to be a father figure.
>
> Pollux: Polonius is a grand father figure.
>
> where -- in this context -- it is altogether ambiguous whether Pollux means 'an excellent father-figure' or 'the figure of a grandfather'. [p. 25]

Having discarded the notion of stress as "perceived relative prominence" of syllables and understanding it to mean instead "potential for receiving pitch accent" -- following Bolinger -- and strictly separating this potential for pitch accent from the manifestation of pitch accent, it is possible to assess the true function of all the phenomena that were traditionally lumped under the title of "stress$_{ch}$." The manifestation of pitch accent on the stressed$_b$ syllables of the words in a sentence indicates the relative importance or informativeness of these words -- from the point of view of the speaker. This can hardly be predicted by

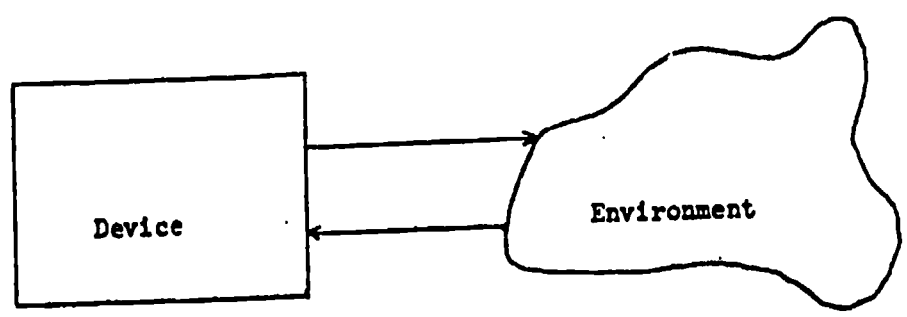
the listener; he has to pay attention to the acoustic signal, in particular to the pitch variations. Furthermore the speaker cannot neglect making these particular placements of pitch contours. Also, if the potential for pitch accent is realized for a particular lexical item this serves to identify the word in the same way as the component vowels and consonants do. In a particularly noisy situation the placement of the pitch contour could help to differentiate "capitalize" from "capitulate", for example. In such cases, again, the listener would have to pay attention to the physical properties of the acoustic signal. However in most cases the listener may be able to identify words without using or hearing the placement of the pitch contours -- indeed, in sentences where some words are de-accented the listener has no pitch contour to go by. And yet he does of course know which syllable in a given word is the stressed_b syllable. Is this done by analysis-by-synthesis? No. At least not if we accept the common characterization of an analysis-by-synthesis system, namely, that it "... involves the internal synthesis of patterns according to certain rules and a matching of these internally generated patterns against the pattern under analysis" (Stevens and Halle 1967, p. 88). If the pattern under analysis contains no overt indication of which syllable has stress_b, then no amount of internal pattern generation is going to help. None of the internally-generated patterns is going to show any better or worse match with the incoming signal than any other. The listener knows which syllable is stressed_b not by using an analysis-by-synthesis procedure but by referring to stored knowledge of the language. In much the same way we can internally generate the first name "Charles" (and much other information) if we hear "Dickens, author of *Great Expectations*." Whether, in the case of assigning one level of stress_b to words, this stored knowledge consists of a large dictionary or a set of generative rules is, as indicated above, entirely an empirical matter. This strikes me as an extremely interesting question well worth the time and imagination the generative phonologists or anyone else might spend trying to answer it. The results might even lend substance to the notion that there can be linguistic contributions to the study of mind.

In summary it appears that the pitch variations used in speech are important; they do convey information; they cannot be completely predicted by the listener. An analysis of these pitch variations and the other phonetic details necessary to an utterance help to explain some linguists' impressions regarding the elusive entity of "stress_{ch}."

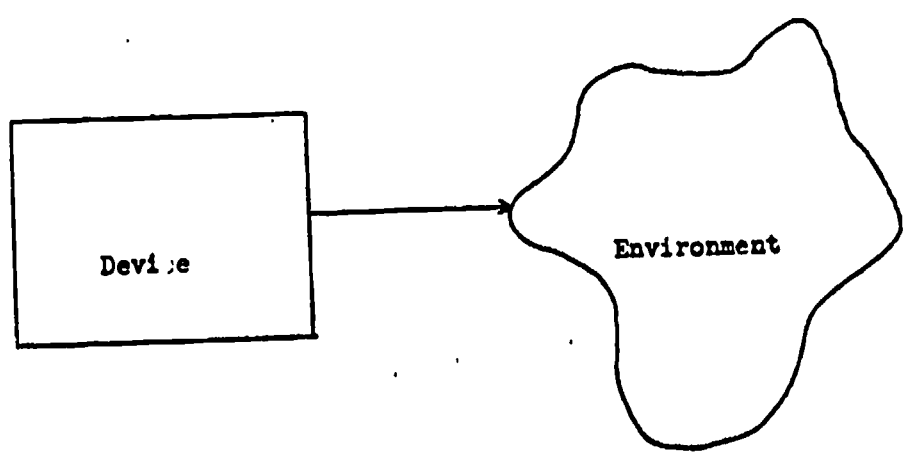
Chapter 3

Prior to discovering the brain mechanisms underlying the observed behavior in speech, it must first be discovered what the brain's task is in speaking. All of the observed actions in speech need not be part of the "program" the brain is trying to execute, just as the minor variations in the shapes of printed letters, e.g., W, W, W, are not part of the brain's task in writing. Given the movements of speech one must first determine which part of them are purposeful and which part of them are fortuitous. A useful place to begin is to make the fairly obvious observation about speech that the order, the timing and the placement of articulators are not random. Some precision in all three of these aspects is essential to communication. These are not wholly independent of one another, e.g., an error in order could be taken as an error in timing, too, but not necessarily vice versa. Still, they can be studied separately. Two tasks for the phonetician, then, are to determine, first, how much precision is needed, and second, how it is achieved. This section will be concerned with preliminary aspects of the second question.

To explain how precision is achieved in these three separate areas it is natural to split the class of possible mechanisms into two: closed-loop versus open-loop, or systems in which the feedback determines the subsequent characteristics of the output, versus systems in which feedback has no such effect and the characteristics of the output are determined by a prior plan. Common analogues for these two systems are automatic pilots in airplanes, which are closed loop, and traffic lights, which, unfortunately, are too often open loop. These two systems are usually schematized more or less as in Figure 44. Various writers have considered the possibility of both these systems being used in speech but have not always considered them for all three of the above mentioned aspects of speech: order, timing, and articulation. Thus, it seems that Lashley (1951), Fairbanks (1954), and Lenneberg (1967) were primarily interested in accounting for the correct ordering of events in speech (although



"Closed-loop System"



"Open-loop System"

Figure 44.

Fairbanks did consider aspects of timing, too); Ladefoged (1967a), Ladefoged and Fromkin (1968), Fromkin (1965 and 1968) and MacNeilage (1968) gave their attention to how speakers achieve target positions in articulation, whereas Chistovich and her associates (1965) considered all three aspects in various places in their work. Bernstein (1935 and 1967), although concerned with all motor skills in general rather than with speech in particular, provided some brilliant insights into the possible role of feedback in determining all of the various qualities of motor output in organisms.

Besides some difference in the problems considered, there is also some minor non-uniformity in terminology and in what is implied by the diagrams used to schematize the two systems. Some of the various diagrams and terms used to describe them are given in Figure 45. For the most part these differences are not significantly different from what has been called above closed-loop and open-loop with the exception that the various kinds of "chain" models may be taken as implying a system which has discrete units in the output signal, and in which the successive units were triggered off by the accomplishment of the preceding units in an all-or-none fashion. One need not accept these limitations. It is possible and more plausible that there is -- if only at a low level in the neuromuscular apparatus that controls movement -- an analog system in which efferent (motor) signals are continuously fed to the muscles and the afferent (sensory) signals are continuously sent back to the brain, and further, in which the information supplied by the afferent signals is used to regulate the characteristics of the output signals over a continuous range.

However, the appropriateness of the closed-loop system for speech has been questioned by Lashley (1951), Cooper (1967), Lenneberg (1967) and by Chistovich et al. (1965)*. Therefore prior to considering the experimental evidence on the possible use of feedback in speech, these theoretical objections should be examined. Of course it must be acknowledged that in ruling out or arguing against closed-loop systems these writers were primarily concerned with the model's ability to account for only one phenomenon, that is the ordering of gestures, and their

---

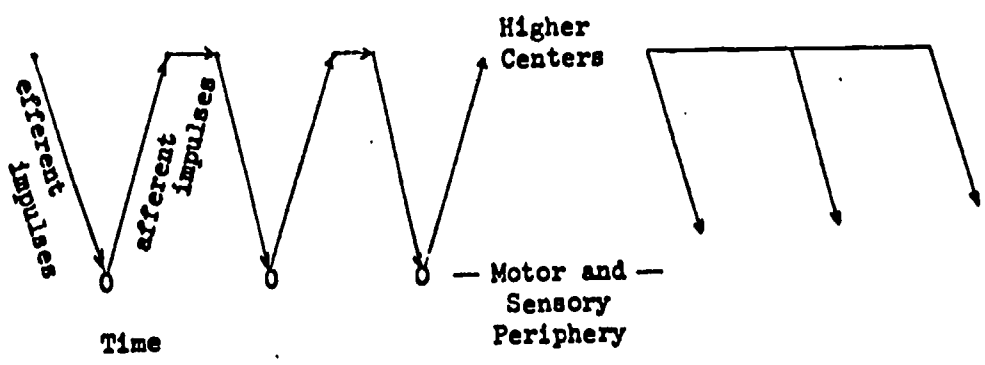
* However, since Chistovich and her colleagues endorse a closed-loop system later on in their work their objection to it was perhaps for the sake of argument.

"Chain" Hypothesis "Comb" Hypothesis

Figure 45a. From Bernstein 1935, 1967

"Sequential Chain Model" "Plan Model"

Figure 45b. From Lenneberg 1967



"Hypothesis 1" "Hypothesis 2"

Figure 45c. From Chistovich, et al. 1965

Figure 45. Various diagrams and terms used by several writers in discussions of open-loop versus closed-loop methods of controlling sequencing (ordering) and timing of gestures in skilled behavior.

arguments need not apply (nor, as far as can be understood, were they intended to apply) to other phenomena, whether ordering, timing, or articulation. In the next section, then, the various arguments leveled against the closed loop, or "chain" model in speech are reviewed and discussed.

ARGUMENTS AGAINST CLOSED-LOOP SYSTEMS IN SPEECH.*

There are at least five such arguments.

1. There is not enough time for a round-trip transit of an afferent then efferent nerve impulse. Lashley (1951) presented this argument primarily against the possibility of a chain of reflexes being able to account for the rapid skilled movements of the hand and fingers in such tasks as playing the piano, etc. But by implication he extended this to all rapid skilled movements. He noted, as did Chistovich and her colleagues, that human motor responses to sensory stimuli (including kinesthetic or deep muscle sensation) typically took as long as 1/8 sec., a much larger interval than the shortest intervals observed in skilled behavior including speech. This argument seems to have been endorsed by Cooper (1967) as well.

2. The "chain" model cannot explain the great permutability of the units that occurs in speech. If it were the case, Lashley argues, that the gestures for the "p" in "apt" triggered the gestures for the following "t", then every occurrence of "p" ought to be followed by "t", but clearly this is not the case. This argument, which was endorsed by Lenneberg, applies as well to any level of language, from the most elementary articulations to syllables, words, phrases, and on up to sentences.

3. Lenneberg claims that the anticipation or pre-planning such as is evident in speech cannot be explained by the chain model. In words such as "stew" and "stay" the lips and parts of the tongue are already partly in position for the vowels even while the consonants are being executed. Moll and Daniloff

* Portions of the following section have appeared before in Ohala and Hirano (1967) but have been substantially revised here.

(1968) have shown that coarticulatory behavior can span syllable and word boundaries, e.g., lowering of the velum begins simultaneously with the tongue movement for the first vowel in utterances such as "freon" or "free Ontario." If one assumes that phonemes are the stored units which are linked together in speech, a model which proposes that the accomplishment of one phoneme triggers the start of the next cannot explain why articulatory aspects of more than one phoneme can occur simultaneously.

Related to this is the observation by Lashley that errors in speech, e.g., spoonerisms, are frequently of a anticipatory nature -- that is, "jumping the gun" -- suggesting that all the elements of an utterance are "in a state of partial excitation" before they are uttered. This, too, argues against a simple chain model whereby a unit is excited or triggered by the execution of the preceding unit.

4. As a corollary to the first argument, above, Lenneberg further argues that owing to the differences in nerve impulse conduction times between the brainstem and distant parts of the vocal apparatus, coordinated simultaneous gestures between, say, the lips and larynx would require that the commands for the larynx be issued some 30 msec. before the commands to the lips. Thus, the representation of the sequential muscular commands in the brain would have to be completely disorganized in the time dimension with respect to action at the periphery, i.e., in the muscles themselves. A single stimulus could not trigger a complex of commands which had to be temporally staggered.

5. Another theoretical objection, which, it seems, has not been made previously, is that the chain model would make it difficult to account for changes in the rate of speaking. In its simplest form the model would predict that after the initial movement, successive movements may be delayed only by the amount of time it takes the appropriate nerve impulses to travel from the sensory receptors to the brain and then back to muscles, since the rate of conduction of nerve impulses is fixed and cannot be voluntarily altered. Yet speakers can say the same word and the same sentence at different rates.

However, in spite of these arguments a case can be made for a modified version of a chain model to be operative in speech -- that is, a case which is no more speculative than those objections raised against it. To this end, it shall be argued that, of the above objections to a chain model, the first and fourth are unfounded and the remaining three require a refinement and limitation of the model.

The possibility that there is not enough time for a nerve impulse to make a complete round trip from sensory receptors to the brain and back again to the muscle in the time speech gestures are observed to follow one another is not so obviously true. Kugelberg (1952) and Rushworth (1966) have demonstrated the existence of facial reflexes in humans, such as the eyeblink and the masseter reflex (the reflex which resists external non-voluntary displacement of the jaw) which have remarkably short latencies: of the order of 12-15 msec. and less. Being basic reflexes it is true that the delays are as short as they are because they involve very few intervening synapses in the reflex arc, and it is perhaps doubtful that this could be true of speech gestures, too. However, as Fairbanks (1954) noted, it is not necessary that a gesture has to be 100% completed, i.e., that complete closure be made between the lips or the tongue and palate, before afferent information on the progress of the gesture is reported to the brain. The brain is capable of predicting the closure on the basis of preliminary information. It also seems to be true that as sequences of movements become more and more skilled, and speech certainly is, control of the movements is more and more removed from conscious control and is given over to more or less automatic relfex levels of control. Thus it is not proper to take 1/8 sec. as the limits of human reaction time. In addition, the fact that the delay between the brainstem and the laryngeal and respiratory muscles is greater is not pertinent here because the most rapid sequential movements in speech are limited to the oro-facial region.

Also, as Lenneberg noted, if one argues instead that these sequential chains are accomplished entirely in the brain itself, the conduction times are negligible and cease to be a problem. However, he says this alternative is neurologically naive (p. 99). It certainly would be if someone actually expected a completely new nervous circuit to be formed in order for a particular chain of motor actions to be formed. But surely it is possible to exploit existing neural circuitry for this purpose. It would seem that the brain is capable of forming new associations and connections between ideas and actions and words all the time and although no one knows how it is done no one denies that it is possible merely because one suggested explanation has been shown to be untenable. Certainly such connections between units could be programmed anew for each speech act and then erased when no longer needed.

These arguments are tentative; much more research is clearly needed. However it seems that the argument that there is not enough time for gestures in speech to be triggered or released

by afferent impulsation from preceding gestures is based on evidence that is not unshakable.

Lenneberg's corollary objection to the "chain" model, the fourth one listed above, has two parts to it: (a) there are sequential gestures in speech which require precise coordination (±5-10 msec.*) between distant articulators, i.e., the larynx and the articulators in the oro-facial region, and (b) the differences in the conduction times of nerve impulses to various muscles used in speech may be as much as 30 msec. There is some evidence that such precise coordination is required between articulators in the oro-facial region, but what is the evidence that such coordination is present between those articulators and the larynx? As far as can be determined there is none. This claim is based on, first, an incredibly bad and amateurish "experiment" which one gathers Lenneberg actually ran, and, second, on a misapplication of the results of some of the work of others on listeners' responses to synthetic and altered natural speech.

In order to show that "there must be considerable precision in timing if laryngeal mechanisms are to be integrated with oral ones" (p. 97) Lenneberg gives some measurements (in msec.) derived from some spectrograms of the two words "obtain" and "optimal" as follows:

Word	First Vowel	Duration of labial stop [sic]	Duration of the aspiration
Obtain	90 [msec.]	170	50
Optimal	110	160	20

[p. 96]

* There is no specific statement of how precise this coordination has to be, only that there must be "an accuracy of milliseconds" (p. 120); but in that it should be "well below 20 msec." (p.97) this probably means a precision of ±5 msec.

These figures are alleged to show that " the acoustic cues for the distinction of certain voiced and unvoiced stops are the duration of the preceding vowel and the duration of the silence during which the lips are closed." (p. 97). In this case the vowels differ by only 20 msec. and the labial stops are said to differ by 10 msec. Also, in the second word "the duration of the aspiration (the only part of the /t/ we hear [sic]) lasts only 20 msec." (p. 97). It is indeed true as Lenneberg indicates that the voiced-voiceless distinction of consonants may be signaled by the differences in the duration of the preceding vowel or the consonantal closure itself (House and Fairbanks 1953, Denes 1955, Lisker 1957, Peterson and Lehiste 1960, Fintoft 1961, Delattre 1962, Fischer-Jørgensen 1964). However all previous studies have found that, other things being equal, it is always the vowel before the voiced consonant that is longer and, conversely, it is voiced consonants which are shorter intervocallically; just the opposite tendency from the one that Lenneberg's figures show. Clearly some other thing -- the phonetic environment -- is not equal in these two words. The accent falls on opposite syllables: obtaín, but óptimal; and this causes vowel reduction -- both in quality and quantity -- in the first word and is sufficient to render the vowel before the voiced consonant shorter than the vowel before the voiceless consonant. Further it is not possible to determine the duration of the labial closure from a spectrogram with words like "obtain" and "optimal". The silence between the vowels is due to the closures of both the bilabial and the alveolar stops. Part of the silence is indeed a good cue for the /t/; the aspiration is not all that is used to identify it. Normally the increased duration of vowels before consonants -- in accented syllables -- is about 60 to 100 msec. depending on the speaker and how fast he is talking. And the increased duration of the voiceless consonants over voiced consonants between vowels and after accented syllables is about 45 msec. (Lisker 1957). It is not clear what significance Lenneberg attaches to the fact that the aspiration of the /t/ in "optimal" came out to be 20 msec. It certainly does not imply any kind of coordination between articulators nor does it imply the existence of an articulatory event of the same short duration. However to make this point clear a brief discussion of aspiration is in order.

* * * *

Aspiration is the delay in onset of voicing after the release of the constriction of an oral stop (Jones 1962, Abercrombie 1967). Voice onset time (VOT after Lisker and Abramson) with respect to the oral release in English and other languages has been measured extensively and accurately by Lisker and Abramson (1964). Their data

revealed that the VOT for initial voiceless stops in accented syllables in English is not very precise: it varies in a normal Gaussian manner over a range of 20 to 120 msec. for /p/ with an average value of 58 msec. In the case of the English /b/, however there was very little variation in the VOT, 75% of the tokens having a VOT of 5 msec. But this does not necessarily imply articulatory coordination between the lips and larynx -- for the following reasons. Two factors determine when voicing may begin: the right combination of the degree of adduction of the vocal cords and the amount of pressure drop across the vocal cords.

From photoelectric glottographic studies of the changes in glottal aperture it seems that for an initial "voiceless" stop the vocal cords are thrown rapidly open and then are brought towards one another but are not adducted completely before the oral constriction is released. The delay in voicing -- causing the "aspiration" -- is due to the time it takes the vocal cords to approximate sufficiently. Variations in the relative timing of the laryngeal and oral gestures, variations in the air flow through the glottis, and other factors contribute to the large fluctuations in the VOT for English voiceless stops. In the case of the initial "voiced" stops and those stops that occur after /s/, however, it appears that the vocal cords are already in the proper adducted position before the oral constriction is released and all that is necessary for voicing to begin is that there be a sufficient pressure drop across the glottis. This is accomplished as soon as (or at most 5 or 10 msec. after) the oral constriction is released. There is little delay and little variation in the VOT with respect to the release of the oral constriction in this case because they are mechanically linked.

Along these same lines it is interesting that in Lisker and Abramson's data "fully-voiced stops", i.e., stops in which the voicing begins well before the oral release (e.g., French /b/), also had considerable variation in the VOT just as did the voiceless "aspirated" stops. Again, this is evidence of the lack of precise coordination between the oral articulators and the larynx.

In a later, more detailed study on VOT in English, Lisker and Abramson (1967) demonstrated that the difference between the VOT's for the voiced and voiceless stops is not precisely maintained in running speech, especially in unstressed environments.

* * * *

Lenneberg claims that experiments at Haskins Laboratories have shown that temporal differences with magnitudes "well below 20 msec. are of the essence" for differentiating speech sounds (p. 97) and refers the reader to Liberman, Delattre, and Cooper (1952), Schatz (1954) and Liberman, Delattre, Gerstman, and Cooper (1956). The first two articles are quite excellent studies dealing with synthetic and natural speech, respectively, and demonstrate the interaction of the stop burst frequency and the following vowel in influencing listener's identification of the place of articulation of the stop. It is not at all clear what relevance these studies have to the point Lenneberg attempts to make, unless it be to note the VOT used in the stimuli, which, as has been pointed out above, offers no evidence of the precise coordination Lenneberg claims is necessary. In Liberman, et at. (1956) it is demonstrated that by varying only the duration of formant transitions a certain stimulus could be made to change from /b/ to /w/ to /u/ and another stimulus could undergo the transformation: /g/ to /j/ to /i/. In no case could it be concluded that temporal differences of well below 20 msec. made any great difference. At best a difference of about 30 msec. in the duration of the formant transition would change the subjects' identification of the stimulus from /b/ to /w/. But the relevance of this to Lenneberg's point is still far from obvious. First of all, these different durations of formant transitions would, in natural speech, be controlled by one articulator, not two, and would therefore offer no evidence of the necessity of precise coordination between distant articulators. Second, these experiments only show what the limits of human discrimination are; they cannot and do not offer any evidence as to the lower limits of the speed and precision with which the articulators actually move in speech. What evidence this study does offer on the range of the speed of formant transitions which the articulators must control in order to maintain the differences in the various "phonemes" suggests that a speaker has a lot of lee-way in the control of his vocal organs.

Further inconclusive evidence offered for this point by Lenneberg are the examples of aphasiac's methatheses. Saying /taks/ instead of /task/ may quite easily be an error of order not of coordination -- at least insofar as this transcription represents the phonetic facts*, that is, a correct pronunciation of /taks/

---

* It would have been helpful if an I.P.A. transcription had been used in these cases; what is one to make of the statement on page 97 that one aphasiac replaced *is* by *si*?

takes about as much coordination as that required to say /task/.

It would appear that Lenneberg has presented no evidence that precise coordination (±5 or 10 msec.) between distant articulators is necessary or present in speech. However, what about the evidence for the other half of Lenneberg's innovative claim: that there may be a difference of up to 30 msec. in the transit time of nerve impulses from the brain to the larynx and the brain to the orofacial muscles?

Considering the amount of space Lenneberg devotes to conclusions based on this claim and the lack of qualification with which these conclusions are stated (p. 103, 120), he is quite cavalier in the matter of providing the reader with any details as to how he arrived at the precise figure of 30 msec. After citing some of Krmpotic's measurements on the length and average diameter of the various nerves which supply muscles used in speech he says only

> Since there is still some uncertainty about the physiological interpretation of these determinations, we need not be concerned here with the details. Suffice it to say that the anatomy of the nerves suggests that innervation time for intrinsic laryngeal muscles may easily be up to 30 msec. longer than innervation time for muscles in and around the oral cavity. [p. 96]

Short of actually stimulating someone's midbrain and measuring the time it takes for the nerve impulse to travel to various muscles used in speech, the transit times of nerve impulses can only be estimated very roughly using indirect information in the literature. In fact it is more accurate to say that the uncertainty surrounding the issue of transit time of a nerve impulse from brain to muscle in the cranial nerves prevents us from making any estimate whatsoever. Lenneberg apparently conducted no experiment on this problem, nor provided any details on how the value of 30 msec. was computed, but some attempt can be made to reconstruct the calculations here.

It is well established that the velocity (V) of a nerve impulse varies directly with the diameter (D) of the nerve fiber*,

---

* That is, other things being equal, such as the presence of a myelin sheath around the nerve. This is true of most of the fibers in the cranial nerves which are the only ones that are considered in this discussion.

for i.e.,

$$V \sim D, \text{ or} \qquad (1)$$

$$V = kD \qquad (2)$$

with V in meter/sec. and D in microns (μ). In order to compute how long it will take a nerve impulse to travel from the midbrain to a muscles in the larynx, one needs to know (1) the diameter of the nerve fiber it will travel over, (2) the length of the nerve fiber, (L), in meters, and (3) the value of the constant k. Thus, the total time of transit of the nerve impulse will be

$$\underset{\text{msec.}}{T(\text{ime})} = \frac{L}{V} = \frac{L}{kD} \qquad (3)$$

At this point there are some uncertainties. The muscles communicate with the brain via nerve trunks, each consisting of several thousand individual nerve fibers which have different diameters and different functions. Some nerve fibers, the alpha fibers, innervate the muscle fibers directly, some, the gamma fibers, innervate the intrafusal muscle fibers which are associated with the spindles and thus indirectly affect the extrafusal muscle fibers, and some are purely afferent, i.e, sensory fibers. It is the first group, the so-called alpha fibers, that are of interest for they would give the lower limit of the transit time of a nerve impulse from the brain to a muscle. But, given a nerve-diameter frequency histogram, it is no simple matter when dealing with human subjects to determine which fibers are alpha fibers. The alpha fibers are very large so it would seem appropriate to use the maximum fiber diameters in the computations. This will be done here, and for the sake of argument and balance use will also be made of the mean values for the fiber diameters. Another uncertainty is the value to use for the constant k. This has been calculated as being 6.0 (Hurst 1939*) from experimental data obtained with cats' alpha motoneurones -- and perhaps it is the same for human cranial nerves.

For the sake of argument, one can take the values for L and D as given by Krmpotic (1959) as being accurate although she reports a range of nerve fiber diameters in the recurrent laryngeal nerve which is significantly more restricted and which yields much smaller mean values than those reported in Faaborg-Andersen

---

* Although compare Boyd's (1965) values of 5.6 and 5.9.

(1957) and Scheuer (1964). It should be revealing to calculate the greatest differences in transit time of nerve impulses between two nerves supplying different muscles in the orofacial region and between two nerves supplying the laryngeal muscles on the one hand and one set of muscles in the orofacial region on the other. The relevant values L and D (range and mean) for three nerve-muscle sets are given in Table IV (from Krmpotic 1959).

Table IV. Values for Length of Three Nerve Trunks and the Diameters of Their Component Nerve Fibers

Nerve Trunk	Muscle Supplied	Length of Nerve Trunk	Range of Diameters of Nerve Fibers	Mean Diameter
Branch of Trigeminal	Tensor velii palatini	4.7 cm.	6-18 μ	9.23 μ
Branch of Facial	Orbicularis oris	26.38	8-15	10.34
Recurrent Laryngeal Nerve	Interary-tenoid	32.2	1-9	5.4

From these figures and by using the value of 6.0 for k in equation 3, above, one can derive values for the transit times of nerve impulses to the three muscles as in Table V.

Table V. Transit Times of Nerve Impulses to Various Muscles

Value for D	Time to Larynx	Time to Lips	Time to Velum
Maximum D used	6.0 msec.*	2.9 msec.	.5 msec.
Mean D used	9.9 msec.	4.2 msec.	.8 msec.

* Flisberg and Lindholm (1970) via direct measurements found the nerve impulse velocity in the recurrent nerve at room temperature to be about 56 meters/sec, yeilding a nerve impulse transit time to the larynx of 5.7 msec. This agrees remarkably well with my calculations, done independently.

From these values one can get the maximum differences in transit time between two different nerves in the orofacial region and two different nerves supplying the larynx and a muscle in the velum. These represent the worst cases of lag.

Table VI. Maximum Differences in Nerve Transit Times for Two Pairs of Nerves

Value for D	ΔT = Time to Lips - Time to Velum	ΔT = Time to Larynx Time to Velum
Maximum D	3.1 msec.	5.5 msec.
Mean D	5.7 msec.	9.1 msec.

It is doubtful that a difference of 9.1 msec. -- if indeed it is that large (the smaller values for nerve fiber diameters of the recurrent laryngeal nerve as given by Krmpotic were used rather than larger values given by other investigators) -- will present any great problem for the kind of coordination required and observed to exist between articulators. Further, these represent, in a sense, worst values; all other groups of muscles and the nerves supplying them will have smaller differences. It is not known how Lenneberg arrived at the value of 30 msec.* However, pending the detailed publication of what this figure means, one can only conclude that this particular objection to the "chain" or closed-loop model is as yet unsupported by any convincing evidence. However, even if this argument must be rejected it is still desirable that we try to obtain a careful experimental determination of the lower limits of coordination between the various articulators. This is just the type of basic data that

* Perhaps this value was arrived at by considering that movement can only be initiated by first going through the gamma efferent system (Hunt and Perl 1960, and Matthews 1964). This may be true but I am not aware that the mechanics of such a system have been quantified sufficiently so as to permit calculations of the time it takes to initiate movement.

is needed in order to begin to build a model of the neurological organization of speech. In this sense it is to be hoped that the issue raised by Lenneberg will continue to receive attention.

All three of the remaining objections have substance, it appears, and argue convincingly against the simplest form of the "chain" model i.e., one which assumes that the various links between succeeding elements are fixed and unchangeable and which are not capable of being "programmed." It is not difficult, though, to modify this model in a reasonable way to eliminate the objections to it. To meet the objection that there is too much permutability in speech to allow a "chain" model, one might begin by observing that in the process of saying something it seems correct to speak of there being a period in which the speaker "decides" what it is he is going to say (cf., e.g., Laver's (1968) model of speech production). In this period he has great freedom to choose and arrange the stored units of speech, limited only by the phonological, syntactic, and semantic constraints of his language and culture. At this stage there is great permutability of the elements of speech. However, after he has selected what it is he chooses to say these units are "partly activated" (as Lashley suggests) and must have a definite order imposed on them. After this point there ceases to be permutability; the order is fixed. After he chooses to say "apt" [æpt] there is a necessity for the tongue closure to be released after the bilabial closure; any other order is wrong and would be counted as an error. Instead of picturing the selected speech units as being "partly activated", it may be conceptually useful to think of them as being put into some kind of "hopper" (or memory buffer) where they are entered in a particular order, the order in which they will be "fed" to the appropriate muscles.

In order to overcome the third objection, that of explaining the anticipation or coarticulation in speech, one might consider that once the stored units are put into our conceptual hopper they can undergo a parameter-by-parameter reorganization such as is illustrated in Figure 46, That is, following the rule outlined in the teachings of Daniel Jones : unless otherwise occupied an articulator can begin to assume the position it must next take, i.e., the commands which will eventually result in muscle contractions can be modified to begin at that moment in the list of events where there is no command contradicting it. Henke (1967) incorporates in his model of speech production a "look-ahead" feature which is similar to this. Of course anticipation of this sort can be said to occur only if the stored units of speech are smaller than the domain or time span over which coarticulation

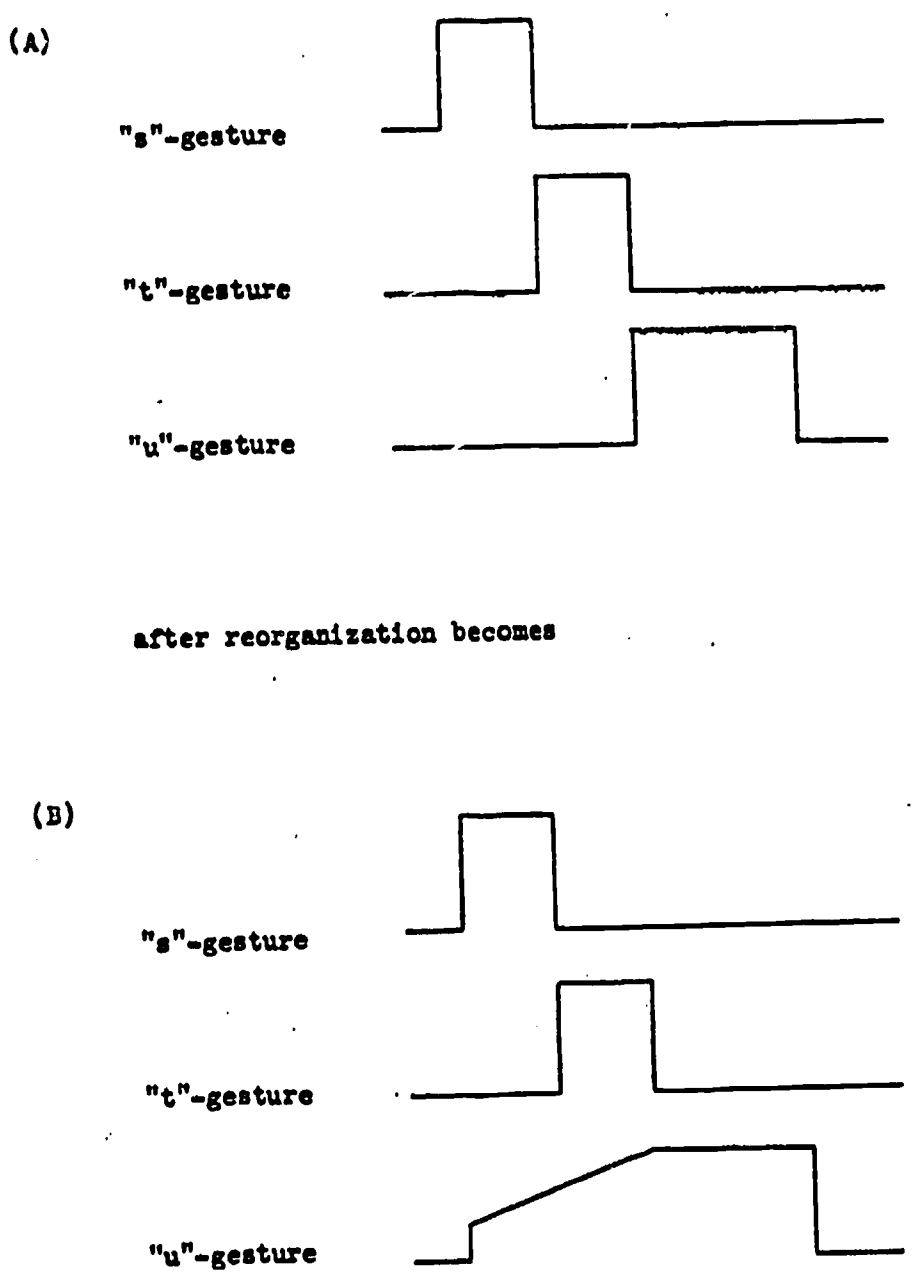

(A)
"s"-gesture
"t"-gesture
"u"-gesture
after reorganization becomes
(B)
"s"-gesture
"t"-gesture
"u"-gesture

Figure 46.

is observed to occur. If we assume that the syllable is the stored unit of speech then the behavior of the lips and tongue in pronouncing the word "stew" is not anticipatory at all. Rather than entering our hypothetical hopper in form (A) in Figure 46, it enters in form (B). Even if this is the case, though, Moll and Daniloff, as noted above, have shown that the domain over which coarticulation is observed to occur may extend across syllable boundaries. Therefore, some such mechanism as has been described is necessary. Now that the appropriate units of speech are inside this "hopper" in the correct order and the onset of the various commands has been parametrically rescheduled, the sequence of chains will not link two "units" as such, but will link the component parts of units. But surely many other solutions to this problem are possible.

Finally, how can the possibility of variations in rate be accounted for with a "chain" model? It is not difficult to imagine how this might be done. There could be variable delay circuits in the brain or possibly variable thresholds on those elements which respond to the afferent impulses and which emit an efferent impulse.

At this point most readers should be objecting that all of this is the merest speculation and tells us nothing about how speech is organized and controlled logically. This is true. This is not the way to do research: instead good experiments are needed. The point here is simply to demonstrate that the armchair speculations which were offered as proof against one possible model for the sequencing of articulatory events in speech can quite easily be countered by other armchair speculations. Such speculative hypotheses are fruitful only if they are married to experimentation; alone, they do little to advance human knowledge.

FEEDBACK IN SPEECH

Is feedback used in speech? Obviously there is some; the question is how much, what kind, and how is it used? Quite clearly speakers do listen to their own voices when they speak, and if they hear a mistake they usually stop and correct themselves. Also, deficiencies or alterations in the auditory feedback channel cause speech to deteriorate, e.g., delaying the auditory feedback by a certain time interval induces stuttering and other disruptions in speaking (Lee 1950, Fairbanks 1955, Mackay 1968), and deafness, if congenital, prevents the development of speech and, if suffered after learning to speak, causes a noticeable gradual degeneration of speech, although, interestingly, it is seldom lost completely.

Of more interest, though, is extremely short-term feedback, that is, information on the immediate progress of speech that can be used by the controlling centers to quickly correct tiny errors (within, say, 10-50 msec.) before they become so big and

so obvious that the speaker has to stop and start again. If present, such rapid feedback would be the task of the various cutaneous and proprioceptive sensory organs in the skin and within the muscles. It is generally assumed that auditory feedback is too slow for this type of "on-line" corrective action, although this may never have been conclusively demonstrated.

Many neurophysiologists and neuroanatomists have testified to the plentiful presence in man of a wide variety of sensory organs in the mucosa and other surface areas of the vocal tract, as well as deep inside the muscles themselves, in particular, the highly useful muscle spindles or stretch-receptors (Hosokawa 1961, Lucas Keene 1961). It is known therefore with some certainty that such feedback is available, and it is known from the studies of Kugelberg and Rushworth, mentioned above, that reflexes using such sensory organs in the vicinity of the orofacial region have very short latencies.

Opinions differ among many speech researchers as to whether or not such short-term feedback is used. In the discussion in a recent conference (House 1967) attended by the most eminent names in the field, arguments and opinions were presented for speech being predominantly closed-loop (using short-term feedback), being predominantly open-loop, and even being capable of switching back and forth between open and closed-loop operation.

But one need not engage in too wild speculation about whether or not short-term feedback from the articulators is used, because there are various bits of evidence that suggest that it is. First of all there are various experiments that have demonstrated the deterioration of speech caused by temporarily eliminating some of this tactile feedback (McCroskey 1958, Ringel and Steer 1963, Ladefoged 1958 and 1967a). Anyone who has taken novocaine at the dentist's usually experiences some slurring of speech due to the temporary loss of sensation in their lips or tongue -- even though the auditory feedback channel is still functioning. If both auditory and tactile feedback are eliminated speech becomes more degenerated than if only one of the feedback channels were eliminated. Still, in listening to tape recordings of speech produced under such circumstances, one is struck by how intelligible it is in spite of the rather severe handicaps imposed on the speakers. (But this may be more a reflection of the great redundancy in speech rather than evidence that individual speech gestures are unaffected by loss of these feedback channels.) In

none of these experiments was it possible to eliminate all relevant sensation from the entire vocal apparatus. Maintaining barely intelligible speech in spite of such sensory deprivation may not indicate how much a given form of feedback is normally used, however. As Ladefoged has suggested, it is possible that under normal circumstances a person may rely primarily on one particular form of feedback but that when that channel is blocked he may be able to shift to another source of feedback.

Fromkin (1965 and 1968) and MacNeilage (1968) in studying EMG records from various articulators discovered that the muscle action patterns of a given articulator for a given consonant or vowel frequently varied depending on the phonetic context in which it appeared. For example, the orbicularis oris, which helps both to protrude or round the lips as well as to close the lips, shows less activity for the rounded vowels /u/ and /o/ when they follow the bilabial stop /b/ than when they follow the non-labial stop /d/ (Fromkin, op. cit.). This could be the result either of the brain storing and executing articulatory programs of about syllable size or of using feedback to determine that the orbicularis oris was already partly activated for the /b/ and thus need not be activated as much for the rounded vowels as it would have to be when they followed the non-labial stop /d/ for which there was no prior activity of the orbicularis oris.

Chistovich and her colleagues presented various interesting pieces of evidence that point to the probable use in speech of short-term feedback in the form of kinesthetic and surface sensation. There was, first, an experiment which was aimed at elucidating the mechanism of initiation and sequencing of the gestures within an interval smaller than a syllable, and involved the measurement of the intervals between two successive but non-contradictory consonantal closures, e.g., -VptV-, -VmnV-, -VkpV-, etc. They found that the minimum interval could be very small -- down to 10-20 msec. but that the two closures never occurred out of sequence, i.e., the "second" closure never took place before the "first" closure, and they never occurred simultaneously. Moreover, histograms of the frequency with which various intervals occurred were positively skewed indicating an apparent "time barrier" which prevented the intervals from becoming too short. As they concluded, the precision represented by these small intervals and the maintaining of the proper order argued against the impulses for the two closures being independent of one another and suggested instead that afferent impulsation concerning the progress of the first gesture was necessary for the reflex triggering or release of the second gesture. (This can also be taken as evidence on the lower limits of how precisely we can

coordinate two articulators. It does not, of course, necessarily mean that in ordinary speech i.e., in a non-experimental situation, we actually *do* or need to maintain such precision between articulators.)

In another experiment of theirs the movement of the lower lip was recorded during the execution of sequences of *labial consonant -- vowel -- labial consonant*. It was noticed that the peak opening of the lower lip varied considerably from one opening to the next, but that the duration of the vowel varied much less. Thus it was found that the velocity of the lower lip on returning to the closed position for the consonant varied directly with the extent of lower lip excursion, i.e., the further it had to go the faster it travelled to get there. Their graph, plotting the closing velocity of the lower lip against the amount of opening of the lower lip is shown in Figure 47. They concluded:

> It is apparent that this phenomenon may be produced by the [regulator] mechanisms ... acting on the lowest levels of control movements. Lowering of the lip (lower jaw) must lead to a reflexive increase in excitability of the centers of antagonistic muscles, while excitability must increase all the more with a greater drop of the lip (lower jaw). [p. 180]

My colleagues and I attempted to replicate this finding except by looking at lower jaw movement alone instead of lip movement. Lower lip movement, as Chistovich and her associates noted, consists of the sum of jaw movement plus any movement contributed by the lip itself. They indicate, of course, that the effect they found may be due to jaw movement. In our study we used a newly-developed device reported on in detail elsewhere (Ohala, Hiki, Hubler and Harshman 1968, and Ohala, Hiki, Hubler and Harshman 1969). The device transduces into an analog voltage the deflection in any given dimension of a tiny light by means of one or more photo-voltaic cells arranged as in Figure 48. The tiny light was attached to a special dental plate for the subject's (JO) lower jaw, this plate having a short stiff wire projecting from it and emerging between the lips.

The analog voltage from the jaw movement transducer was fed to a small computer (LINC-8) which had been programmed to compute the velocity of the movement, find the maximum opening of the jaw and display these and various other related parameters as correlation plots.

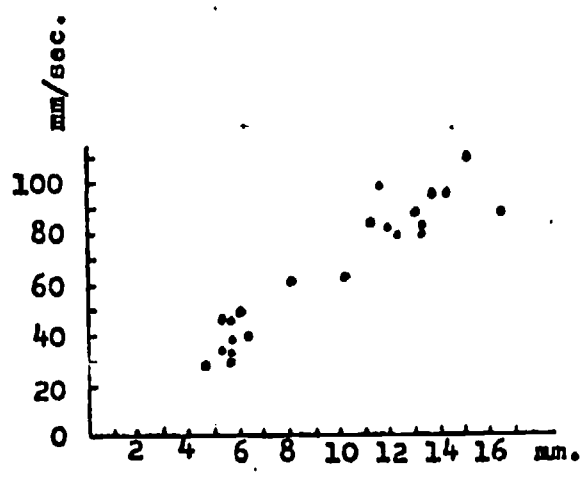


Figure 47. Maximum opening of jaw, plotted on abscissa, against peak velocity of jaw closing, plotted on the ordinate. (Traced from Chistovich, et al. 1965, Figure 5.13)

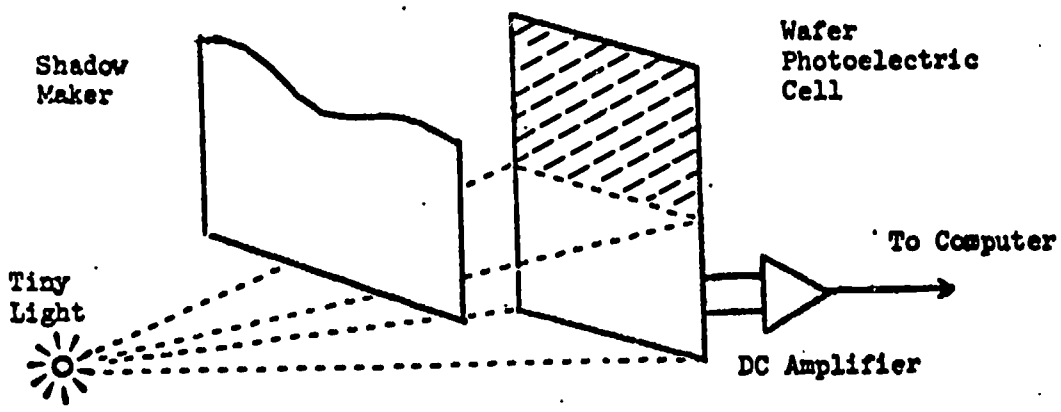

Shadow
Maker
Wafer
Photoelectric
Cell
To Computer
Tiny
Light
DC Amplifier

Figure 48.

The results can be seen in Figures 49a and b which show the plot of maximum jaw velocity (both opening and closing, and thus by convention having negative and positive velocity, respectively) against maximum jaw opening for two words "sas" and "sak" (spoken in isolation). The figures show that both the opening and closing velocity of the jaw is directly proportional to the extent of opening of the jaw. This means the reflex adjustment of the centers responsible for closing the jaw could have been accomplished as early as the moment the jaw began to open or as late as the moment the jaw reached its maximum opening. Further investigation would have to be undertaken to resolve this. The figures show as well that the return velocity is much less for the /k/ than for the /s/. This is undoubtedly related to the fact that the jaw does not need to close as much for velar stops as it does for alveolar-palatal fricatives (cf. records of jaw closure in Ohala, et al. 1968, and Daniloff, Amerman, and Moll 1968). Thus the return velocity of the jaw is not simply an automatic consequence of how far open the jaw is but depends as well on the nature of the following consonantal closure. If this is a reflex action it is capable of being influenced by higher centers of the brain.

This confirms the finding of Chistovich and her colleagues and gives fairly good evidence of the presence and use of short-term feedback to make quick adjustments of articulator movement in speech. But what is the purpose of the feedback? It is possible that it is used to maintain a fairly constant interval of jaw opening. That is, the trajectories of the articulators may be allowed to vary, but the moments of collision may not be. However it is also possible to view it as Chistovich et al. do, namely that the variable velocity of the jaw movement is due to a sort of neuromuscular "spring" effect, such that the greater the displacement of the jaw is, the greater will be the restoring force, and consequently the higher will be the velocity of jaw movement. Since the restoring force is apparently different depending on the nature of the following consonant, one would have to allow that it could be altered by such higher-level non-automatic features. Some marginal evidence for this view, or at least against the view that the effect of jaw velocity varying directly with jaw opening is due to an attempt by the brain to maintain a constant interval between consonants, is the fact that the total interval <u>does</u> increase as the extent of jaw opening increases, as is shown in Figure 50. Further research, involving a greater variety of phonetic contexts and more speakers should determine with more certainty how and why this feedback is used, and whether or not it is used similarly in other articulators.

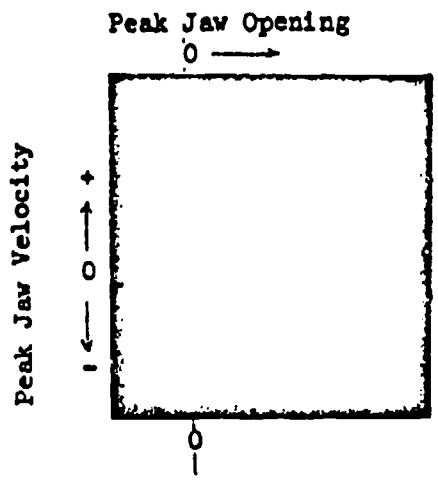


Figure 49a. Fifty tokens of [sas]. Abscissa: peak jaw opening (each division = 2.8 mm.); ordinate: peak jaw velocity (each division = 23.6 mm./sec.). By convention negative velocity pertains to jaw opening and positive velocity to jaw closing. Here, upon closing for the final [s], the peak closing velocity, in mm./sec. equals about 0.5 times the peak jaw opening in mm.

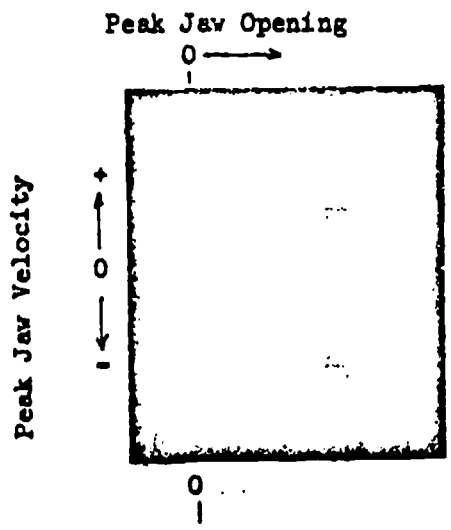


Figure 49b. Fifty tokens of [sak]. Abscissa and ordinate as in Figure 49a. Here when the final consonant is [k], the peak closing velocity in mm./sec. equals about 0.3 times the peak jaw opening in mm.

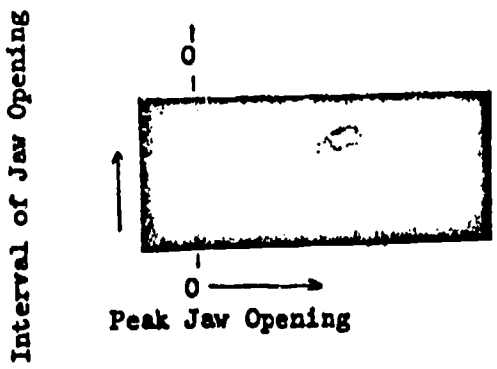


Figure 50. Fifty tokens of [sak]. Abscissa, as in Figure 49a; Ordinate: total open interval (each division equals 38 msec.).

EXPERIMENTS ON THE TIMING OF SPEECH

Certainly one of the high points of current phonetic research is the interesting experiments conducted by the Leningrad group (Chistovich et al. 1965) concerning the timing of events in speech. One of the questions they addressed themselves to was what determined the onset of successive syllables in speech.

In the key experiment they exploited the existence of small variations in the various delays in transmission of neural impulses from place to place. They reasoned that if the initiation of the sequential articulatory units were linked in the manner suggested by the chain or closed-loop model then the variations or fluctuations in the intervals in many tokens of a given sentence ought to accumulate and therefore the variation in the time taken to say the whole utterance ought to be equal to or greater than the sum of the variations of the separate component intervals. If, on the other hand the initiation of the successive articulatory units was dependent on a higher pacing mechanism, and was independent of the progress of the preceding units, then the variation of the whole interval ought to be less than that of the sum of the variances of the component intervals. An attempt was made to gather sentences which minimized possible changes in the rate of speaking. A few hundred tokens of the same sentence spoken by two speakers provided the data and the results showed that the variation of the whole interval was less than the sum of the variations for the smaller component intervals, thus favoring the "comb" or open-loop model for the sequential generation of syllables.

There should be no misconceptions about what the results of this experiment show. Favoring the "comb" model does not necessarily point to the absence of feedback from the periphery and likewise, favoring the other model need not imply the existence of such feedback. In fact, this experiment by itself cannot tell whether or not feedback is used in controlling the timing of units in speech. It would appear that the results could at best help us decide between the following two systems:

A. A "Timing-Dominant" system, i.e., a system which maintains a tight time schedule perhaps at the expense of precise and thorough accomplishment of the gestures.

B. An "Articulation-Dominant" system (for lack of a better term), i.e., a system which maintains precise and thorough performance of the gestures no matter how much time it takes.

Both of these systems could or could not employ feedback. To illustrate this point, consider, for example, the four possible ways a puppeteer could control his puppet. Following Bernstein, the situation can be made more like the one that prevails between the brain and the muscles, and the puppet strings can be made of elastic or rubber. If the puppeteer had to make his puppet march or dance in time to an externally-supplied beat, we would have a timing-dominant system. The puppeteer could attempt to accomplish his task by using visual feedback or by not using it. If he were very, very practised it is conceivable that he might be able to do it without looking at the result of his rhythmic tugging at the strings. On the other hand, since the strings are elastic and the results of a given tug will depend on the degree of stretch of the strings at the moment when it occurs, it is easy to see that visual feedback could be very useful in maintaining the rhythmic step of the puppet. It is also easy to see in this case that in order to maintain a rhythmic step by the puppet in spite of variations in the response of the puppet's limbs due to the elastic strings, a non-rhythmic control of the strings by the puppeteer would be needed. (For further more extensive arguments on this point, see Bernstein 1935 and 1967). At any rate, the observer out in the audience, after timing the movements of the puppet in accord with this experimental procedure would find that he had a timing-dominant system, but he would not, by that experiment alone, be able to tell if the puppeteer were using visual feedback or not. Or, if the puppeteer merely had to make his puppet walk casually across the stage with no absolute time limit, then it would be an articulation-dominant system but again visual feedback could be used or not used. This time the feedback, if used, would aid the puppeteer in making sure the puppet's feet paced the floor one after another, that they touched the ground, and that the puppet stopped before running into the wall, etc. Thus the results of the experiment performed by the Leningrad group, as described above, pointed to the mechanism governing the timing of speech being a timing-dominant system, i.e., one in which a strict time schedule of events is maintained, possibly at the expense of precise articulation. With these qualifications in mind it is possible to return to the use of the terms "comb" and "chain" models.

However, besides the above, there are some other objections that can be raised against the Leningrad group's interpretation of their experimental results. First, one might ask whether in an experimental situation in which the subjects are asked to repeat a sentence over and over again hundreds of times, they may not very soon adopt a rhythm in pronouncing the sentence

whereas ordinarily they would not. Thus the experiment might end up testing whether speech *could* have a fixed underlying time schedule, not whether it usually *does*. Walking over uneven ground is not usually perfectly rhythmic but can be made so in the case of marching. Avoiding this possible experimental artifact would be troublesome but could be done most likely by incorporating other verbal material in the text to be spoken or perhaps by recording the individual tokens of the test sentence hours or days apart. Another hazard of the experiment is that the particular measure used to decide between one model and the other, that of comparing the variance of the largest interval with the sum of the variances of the component intervals, requires that one be sure of the size and number of articulatory units one has in the given sentence. If one miscalculated the number of units which are actually in the sentence it would make no difference if it were a "chain" system, but it could make a difference if it were a "comb" system, especially if one counted too few intervals. In this case one could erroneously accept the "chain" model after finding that the sum of the variances of the (too few) component intervals equalled or were less than the variance of the largest interval. However, since these authors found a "comb" system, this criticism doesn't apply to their results. But a possible miscount of the number of units becomes important again if, instead of a system that is wholly "chain" or "comb", one has a hybrid system, in particular one in which the units themselves are executed in accordance with the "chain" model, but the sequential gestures within a unit are executed according to a strict time schedule, i.e., according to the "comb" model. For example, this would be the case if the words, say, in a sentence were executed one after the other, the onset of each word being dependent on the accomplishment of the preceding word, but if the gestures within the word were executed according to a strict time plan. In this case a miscalculation of the size and number of units when following the procedure outlined by Chistovich et al. could lead to the erroneous rejection of the "chain" model if one counted too many intervals, and, as above, could lead to the erroneous rejection of the "comb" model if too few intervals were counted. This objection does apply to their results.

Chistovich and her colleagues took the units to be syllables based on the results of a previous experiment, in which it was shown that the duration of the words and syllables relative to the duration of the whole utterance remained constant during changes of rate, but the relative durations of the consonants and vowels, the components of the syllable, varied during changes of rate. Thus the smallest interval maintaining its relative temporal "integrity" in the face of changes in rate was the syllable -- at least in Russian. But these results could as well be taken as indicating that the articulatory unit could be no smaller than the syllable but it could be larger. Thus there is some uncertainty surrounding the selection of the intervals measured. Is there any way of doing this experiment without knowing what the articulatory units are? Possibly there is.

The mathematical justification offered by Chistovich et al. for adopting the particular experimental test of comparing the variance of the largest interval with the sum of the variances of the component

From this point to the asterisk on p. 152, the discussion contains serious errors; a revision is in progress.
--JJO, December 1970

intervals seems to be sound, but a simpler procedure seems to give roughly equivalent results, namely seeing if the variance is constant for different-sized intervals, thus favoring the "comb" model, or if it is directly proportional to the mean interval, thus favoring the "chain" model, i.e.,

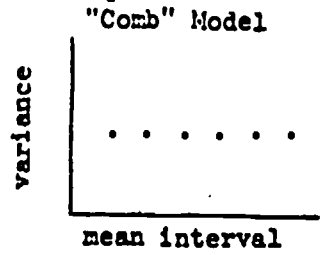
$$V = k \qquad (1)$$

or

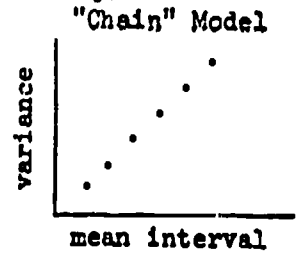
$$V = kI \qquad (2)$$

respectively, where V is the variance of an interval I, and k is a constant. Graphically, this would look as follows:

Equation 1:
"Comb" Model



Equation 2:
"Chain" Model



One would proceed as did Chistovich and her colleagues and obtain the same kind of data, but instead of comparing the sum of the variances of the component intervals with the variance of the largest interval, one would simply see which equation or graph best described the data; there would be no need to know precisely how many units one had in the test sentence. From a sentence in which there are n basic intervals one can extract $\sum_{i=1}^{n} (i)$ intervals in all, e.g., four basic intervals would yield ten intervals and consequently ten data points:

	1	2	3	4
1.	1			
2.		2		
3.			3	
4.				4
5.	1 + 2			
6.		2 + 3		
7.			3 + 4	
8.	1 + 2 + 3			
9.		2 + 3 + 4		
10.	1 + 2 + 3 + 4			

Of course, one must measure only those intervals which one has good reason to believe are the result of neuromuscular signals. Therefore one would not want to measure the intervals between successive taps of the tongue in trills or those between successive vocal cord vibrations because both of these are determined partially by aerodynamic effects not directly under muscular control.

This procedure will work only if the average rate of speaking is the same for all the sentence tokens. This is unlikely, as allowed by Chistovich et al. If there are changes in the rate, it will tend to increase the variance on larger intervals no matter which system is present. That is, the data would be better described by equation 2 in either case. Thus, if the rate of speaking is varied one will not be able to choose between one system or the other on the basis of the results from this procedure.

But there may be a way out of this problem. One may be able to get a collection of sentences which are likely to have been spoken at the same rate if, out of all the sentences recorded, one chooses just those few that have the same or very nearly the same total duration. By thus arbitrarily limiting the total duration of the largest interval one will also lose a few data points since any two intervals which together add up to the total interval will now necessarily have the same variance, since they share a common interval boundary, but are fixed at their other boundaries. That is, the interval that is n% of the total interval will share the same variance with the interval that is (100-n)% of the total interval. Thus one will be limited in effect to looking at the data points for intervals $\leq \frac{1}{2}$ of the total interval.

These procedures may be applied to some of the published data of the Leningrad group. The data points in their Figure 3.13 show that the variance increases with the mean interval duration (cf. Figure 51). The trend of the points cannot be exactly described by equation 2, but would be better approximated by an equation of the form

$$V = kI^c \qquad (3)$$

with $c > 1$.* However, for an approximation to these points one can use equation 2 with $k = 1.04$. They report that out of the original 800 or so sentences they chose for final processing some 400. This was done by graphing the changes in the total duration for all the sentences and then choosing from that graph sentences in a "relatively stationary sector." It is not clear whether or not this procedure means that the final set of sentences were restricted to roughly those having the same total duration and thus that it corresponds to the procedure outlined above. The shape of the graph in Figure 51 would lead one to suspect

---

* Equation (3) is a more general form of equations (1) and (2), above, in which $c = 0$ for the "comb" model, and $c \geq 1$ for the "chain" model.

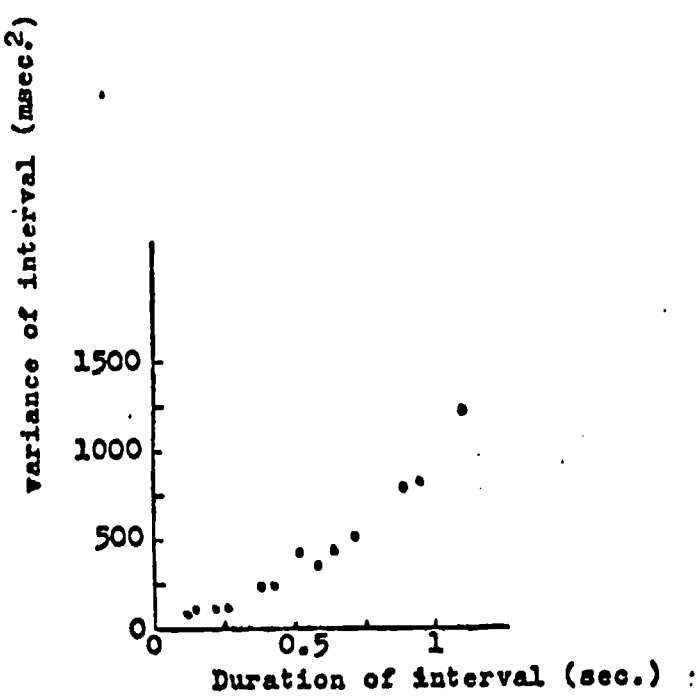


Figure 51. (Traced from Chistovich, et al. 1965, Figure 3.13)

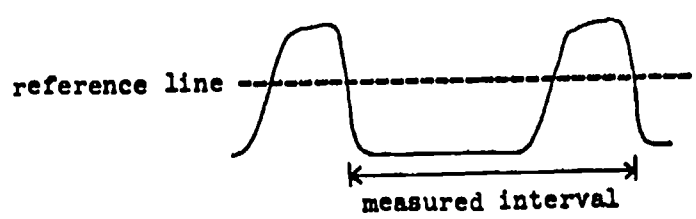
that the sentences did not have a limited range for the total duration. So, if there were changes of rate the results are inconclusive for the reasons presented above. If rate was the same for all the sentence tokens one would have to accept that speech timing was best described by the "chain" model.

I have attempted to perform the experiments and procedures as described above with the exception that no attempt was made to counter the possibility of an unnatural speaking rhythm interfering with the natural timing of the speech. Although they have barely progressed through to the pilot experiment stage, they are worth reporting here.

Such experiments should properly be run using a sizable experimental population -- at least of the order of 20 subjects -- and, since they necessitate much measurement and statistical computations, are most efficient if automated. At this stage though, only one subject, (JO), was used, and the measurement was done by hand. Changes in oral pressure transduced and written on paper by an ink-writing oscillograph, a Siemens Oscillomink, were used as reference points for the measurements. The test sentence was constructed to be capable of easy segmentation. It was:

"Your Paula may put Happy upon your pair of pigs."
↓ 1 ↓ 2 ↓ 3 ↓ 4 ↓ 5 ↓
[yɔrpɔləmeɪpʊthæpɪypɑnyəˈpɛrəpɪgz]

that is, consistently using the reduced forms of "upon" (pɑn) and "of" (ə). This yielded five basic intervals as delimited in the phonetic transcription. Measurements were made from the point the pressure curve fell upon release of the p's and crossed an arbitrary reference line above the "zero pressure" line, as is schematized below.



Measurements were made in millimeters, where 1 mm. = 9.72 msec., and the measurement error is estimated to be ± 0.5 mm. Sixty-four tokens of the test sentence were obtained. The resulting intervals and their variances are given in Table VII. Plotting these variances against the corresponding mean interval duration gives the 15 data points in Figure 52. The trend of points here is more nearly a straight line than are those of Chistovich et al., and equation (2) with a k about 0.83 will

Table VII. Intervals and Variances for the Whole Set of Test Sentences, N = 64

Interval number	1	2	3	4	5
Mean duration (I)	478 msec	330	196	350	298
Variance (V)	325	247	194	149	329

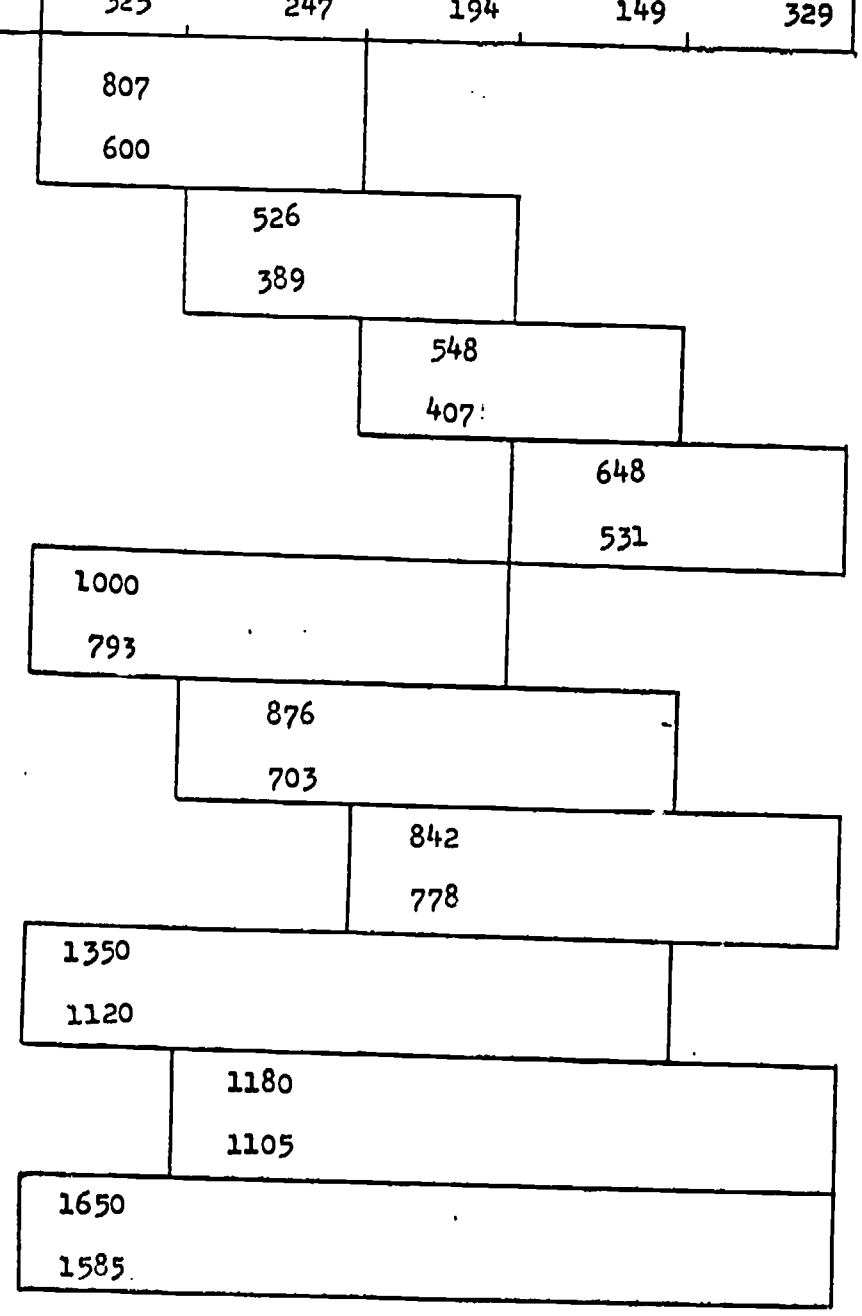
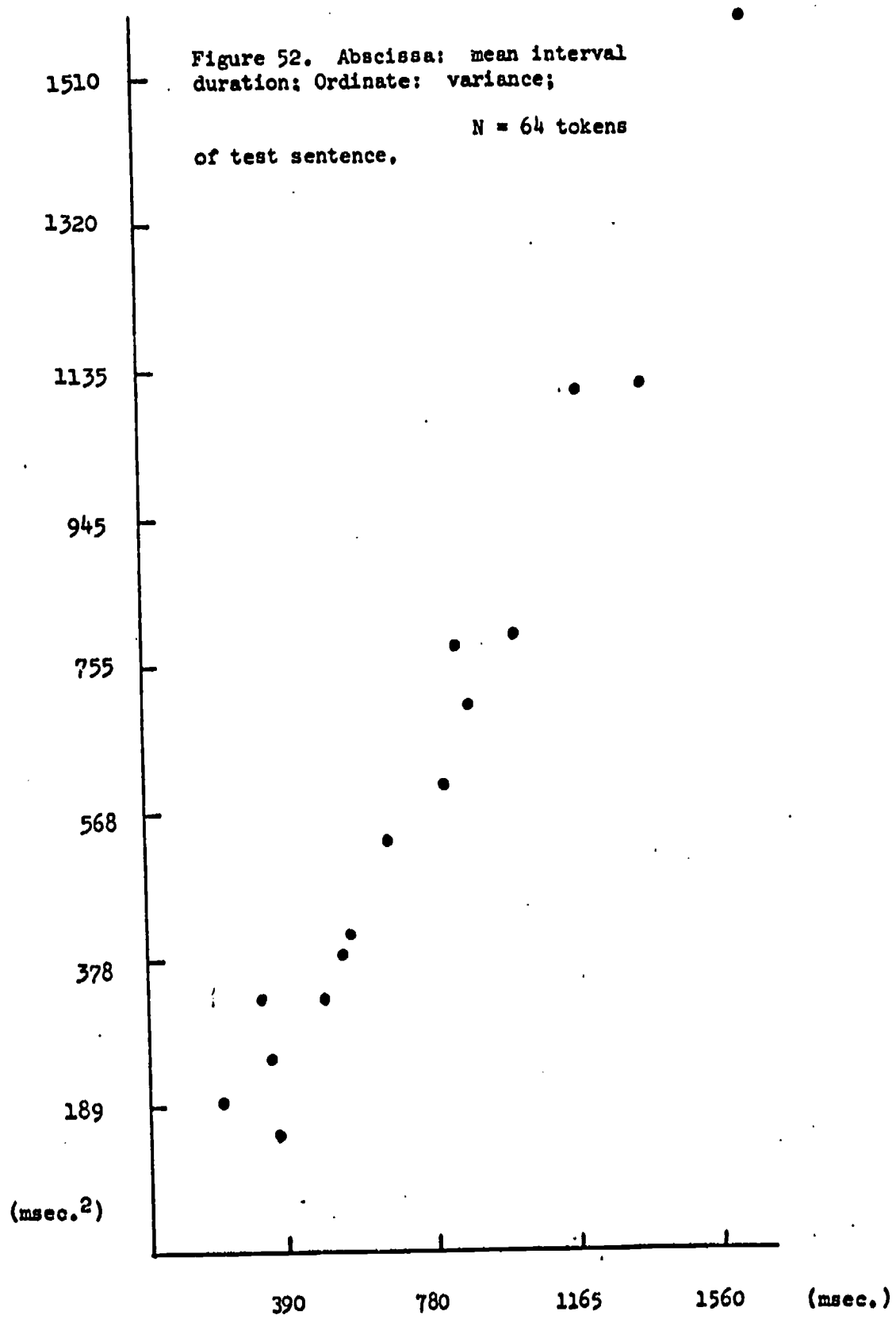


Figure 52. Abscissa: mean interval duration; Ordinate: variance;

N = 64 tokens

of test sentence.



describe the line best fitting these points. If one could be sure there were no changes in rate in the pronunciation of all 64 of these sentences, then one would have evidence that speech was like the "chain" model, according to the arguments presented above. However, one cannot be sure of this, so this graph actually tells very little.

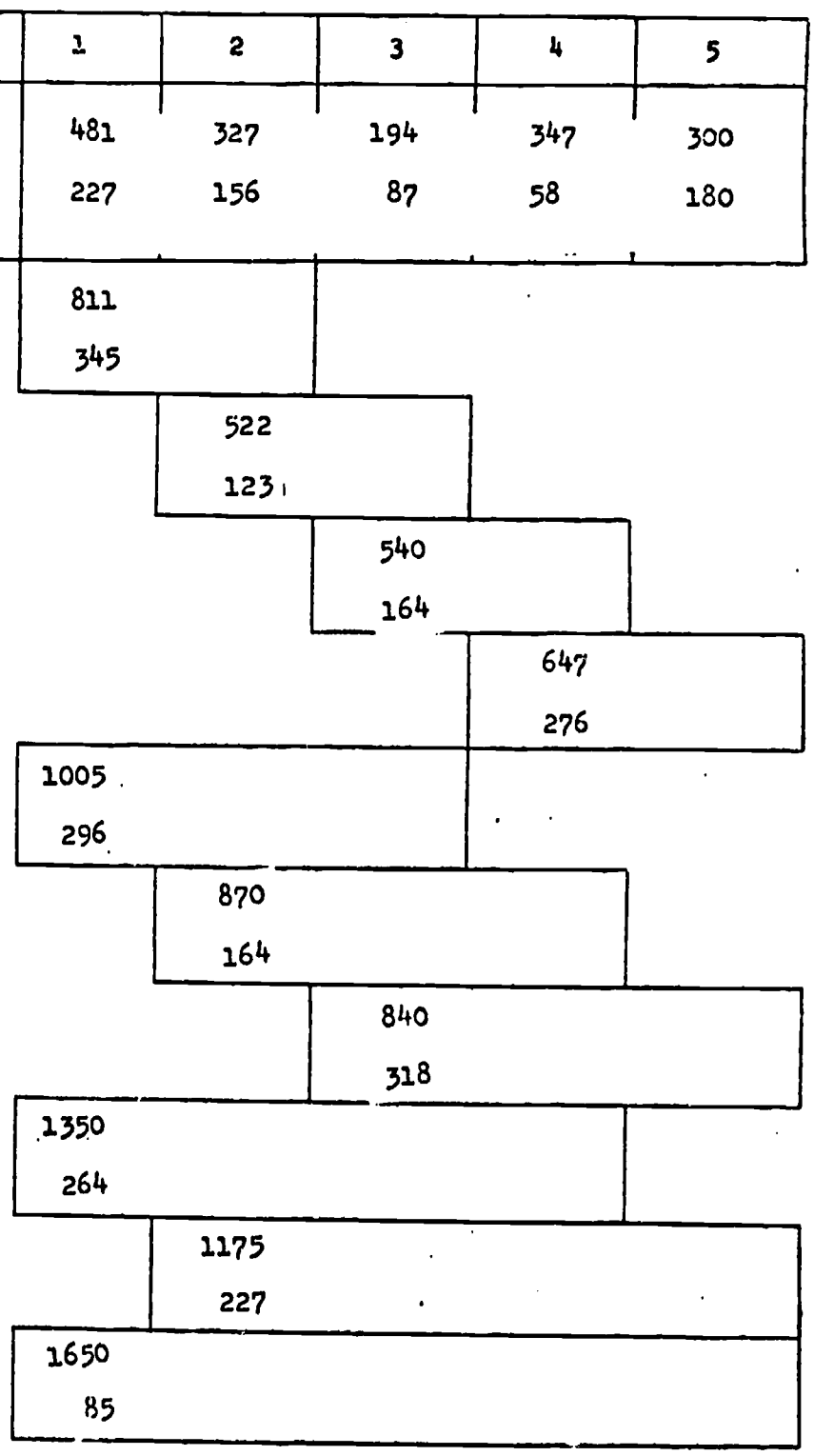
Taking the next step, though, out of these 64 sentences 20 were isolated which had the same total duration ± .15 mm. The resulting values are given in Table VIII. Figure 53 shows the plot of these variances against their mean interval durations (some points being redundant, as mentioned above). From this, although the scatter of the points makes conclusions somewhat shaky, it would appear that the "chain" model must be accepted, that is, that in speech the precise and thorough accomplishment of a speech gesture is desired no matter how much time it takes. However, one situation that would again render this result inconclusive would be if the rate of speaking were not always the same within a given sentence, that is if the instantaneous rate could fluctuate within a sentence. If this were the case it would cause the "bulge" in the middle of the graph in Figure 53, no matter which system pertained in speech. That the instantaneous rate could vary in speech is likely because of the fact that ultimately all time-keeping devices in the universe, including the revolution of the earth around the sun and the most accurate atomic clock, are all what has been described here as "articulation-dominant", that is, a "chain" system. No independent "Platonic" time standard can exist. There are then only clocks of varying accuracy. A time program for speech, if it exists, could keep time by a very inaccurate neuronal clock. In timing speech with a more accurate external hardware clock we would certainly find the system to be like the "chain" model in spite of the possible existence of an independent time program for the sequencing of the speech gestures. At this point more research is needed; until then it may not be possible to conclusively determine how the onsets of the sequential gestures of speech are controlled.

* Chistovich et al. note that the relative error ((Standard Deviation/Mean Interval) x 100) is larger on small intervals than on large intervals, and attach importance to this fact:

> It is significant that with such a large relative error of the duration of the individual sounds of speech [10-20%] the relative error of the duration of the entire syntagma [≡ phrase] amounts to only about 3%. [p. 101]

Table VIII. Intervals and Variances for Subset of Test Sentences, N = 20

Interval number	1	2	3	4	5
Mean duration	481	327	194	347	300
Variance	227	156	87	58	180



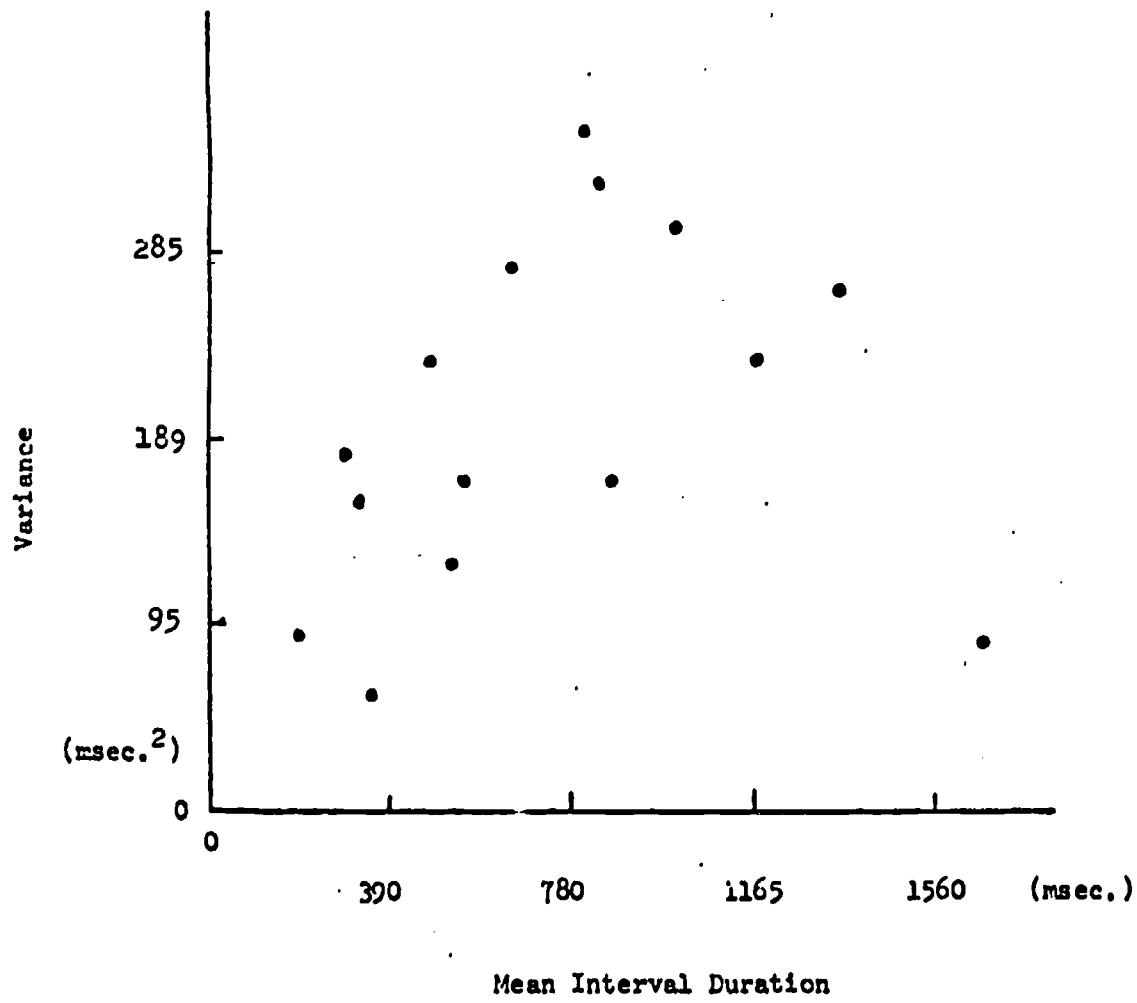


Figure 53. N = 20; the rest is as in Figure 52.

Likewise, Allen (1968) comments on their data and similar findings of his own and notes that

> In order for this reduction in variance to occur, there must be timing information that extends over the whole phrase. [p. 75]

These values for the relative error are undoubtedly true; in the data reported here there were similar values: more than 7% relative error for the smallest interval measured, against 2.4% relative error for the whole interval. However it is not clear how these values imply the existence of timing information independent of the sequence of speech gestures being executed -- that is, that these values point to a timing-dominant system in speech. These values are directly derivable from the equations of the curves in Figure 52 and Figure 3.13 from Chistovich et al., plus the definition of relative error. Relative error, E, as a percentage, is defined as the standard deviation s.d., divided by the mean interval, I, times 100. The standard deviation is, of course, the square root of the variance, V. Using this plus equation (2) above, one obtains

$$E = \frac{s.d.}{I} 100 = \frac{(V)^{1/2}}{I} 100 = \frac{(kI)^{1/2}}{I} 100$$

Taking k = 1.04 for their data, the relative error can be calculated for various-sized intervals as in Table IX.

Table IX. Calculated Relative Error for Three Intervals

Interval	Relative Error
50 msec.	14.4 %
100 msec.	10.2 %
1250 msec.	2.9 %

These calculated values are in agreement with the measured ones, which is not surprising since they are really just slightly different mathematical characterizations of the same basic data. They add no more information to the study and are subject to the same criticisms as are the graphs in Figures 51 and 52. It probably is not advisable to express the fluctuations in these intervals as relative error because it gives the false impression that there is more fluctuation on small intervals than on big intervals therefore indicating that some of the fluctuations are cancelled out or compensated for over longer time intervals. But

this is false and is only a mathematical artefact. The absolute value of a large fraction of a small amount may still be much less than the absolute value of a small fraction of a large amount.

RHYTHM IN SPEECH

One may usefully begin a discussion of rhythm in speech by noting with Allen (1968) that whether speech articulation actually is rhythmic or not, lots of people have the impression that it is. However, the question of the apparent perceptual reality of rhythmicity in speech will not be covered here. Attention will be given here to the rather restricted issue -- and more interesting issue, if it turns out to be true -- of whether or not the successive (measured) intervals between gestures in speech have the same duration or multiples or sub-multiples of the same duration. This would represent a special case of the "comb" model, namely one in which there was not only a separate time program for the events of speech but one in which the primary events occur at equal intervals. We have presented above some evidence that a time program -- whether isochronic or not -- does not appear to exist in speech. However this finding was judged inconclusive and so it may be useful to consider here, using a different approach, the more restricted isochronic "comb" model.

Chistovich and her colleagues (1965) discuss the possibility of an isochronic rhythm generator, but in fact produce evidence that if there is a time program independent from the phonetic material said in time to it, it is not strictly isochronic. Two Russian words which were phonetically identical except that one had an extra consonant, i.e., CVCV versus CVCCV, were each spoken in the beginning of the same sentence frame several times. Although the words differed in duration by about 50 msec., if speech were rhythmic, that is, a system with isochronic intervals between the beginnings of the major units, one would expect that this time difference would be obliterated by the end of the sentence. It turned out that the sentences differed in duration depending on which word it contained. If it had the longer word in it, the sentence itself was longer and by about the same amount as the word. This indicates either that the pacing mechanism is not perfectly rhythmic but in fact takes into consideration the phonetic structure of the individual syllables before working out the schedule for the entire utterance or that there is *no* pacing mechanism for a whole utterance.

Similarly, Shen and Peterson (1962) and Allen (1968) were not able to find exact isochronism in English when measuring interstress intervals or even when computing the correlation of durational changes between successive interstress intervals.

Lenneberg (1967) hypothesized that there exists in speech a basic rhythm of 6 ± 1 cps*, which governs the rate of syllable production.

---

* The range is given variously as 6 ± 1 cps., i.e., a period ranging from 143 to 200 msec., or from 140 to 180 msec. (p. 119).

He admits that the neurological basis for this rhythm cannot be guessed at given our imperfect knowledge of the brain, but indicates that it represents a "rhythmic alternation between states ... of initiation and execution of motor patterns (or cycles of activation and inhibition)" (p. 109). To support this hypothesis he gathers together quite an impressive variety of evidence, all of which are said to show that an interval of about 1/6 sec. frequently plays an important role in various speech phenomena. However, one gets the impression that not all of the evidence is equally solid. The evidence on the well-known disruptive effect of delayed auditory feedback assumes the amount of delay producing the greatest interference with speech is 180 msec., but it is clear from the literature that the critical delay varies from subject to subject and generally ranges from 150 to 300 msec. (Chistovich, et al. 1965, Fairbanks 1955, Spuehler 1962, MacKay 1968), a range that does not fit nicely into the limits of 140-200 msec.

Further, the evidence from the rate of syllable production is neither convincing nor accurate. Lenneberg attempts to show that the average rate of syllable production is about 6 syllables/sec. To begin with, by apparently accepting the findings of Stetson as to the nature of the syllable, namely the old "breath pulse" theory (Stetson 1928), Lenneberg seems to ignore the fact that Stetson's findings could not be replicated by Ladefoged et al. (1958) nor more recently by Lieberman et al. (1967). Given that no one knows what the syllable is, one may question the relevance of measures of syllable rate in speech. However, for what it is worth the data that does exist on this doesn't help Lenneberg's hypothesis. He does not accurately report the findings of Hudgins and Stetson (1937) as to what the relative speed of articulatory movements are. First of all, they were concerned with the *maximum* rates at which syllables could be produced with various articulators, whereas one would have thought Lenneberg would be more interested in *average* rates. The range of values he reports they found, 5.5 - 7.5 syllables/sec., is wrong; it should be 5.2 - 9.6 syllables/sec. Table X gives Hudgins and Stetson's published values plus those from a comparable study by Kaiser (1939).

Table X.

Hudgins and Stetson (1937)			Kaiser (1939)			
			Men		Women	
Population	9		116		100	
Articulator	Mean	Range	Mean	Range	Mean	Range
Tip of tongue	8.2	7.2-9.6	7.4	5.6-9.1	7.2	5.3-9.1
Jaw	7.3	5.9-8.4	5.4	4.8-8.3	4.8	4.5-9.1
Back of tongue	7.1	5.4-8.9	6.6	*	6.4	*
Lips	6.7	5.7-7.7	---		---	
Velum	6.7	5.2-7.8	---		---	

*Time base of graph is mislabeled; range cannot be determined.

It is very likely (as suggested by Chistovich et al., p. 135), that the variations are dependent on the articulator used and on that articulator's inertia and not on any central limitation of rhythm. Thus the fastest rate, 9.6 syllables/sec., was associated with the tip of the tongue, and the slowest rate, 5.2, with the velum. Table XI provides additional such evidence from the literature cited by Kaiser (1939) and from the measurements of Chiba (1935). To this may be added my own finding that some subjects are capable of speaking at a rate of more than 10 syllables/sec. in the last part of sentences such as "Dead headed Ed had edited it." In addition, Kaiser (1939) performed studies on the average rate of syllable production for 216 Dutch students speaking Dutch in a variety of speaking tasks: pronouncing lists of words, counting, reading, and spontaneous speech (obtained by having the subjects comment on four pictures). The average duration of the syllables was

Table XI

Source	Population	Language	Value	Meaning
Bourdon (1892)	4	French	8.1 syll/sec	Average rate; reading journal
	1	French	4.3-5.4	Range of average values; reading verse
	1	French	7.4	Very fast rate; reading
Oehrm	?	German	7.2	Average
	?	English	7.4	Average
Richet	?	?	10-12	Maximum rate; reading poetry
Blancquaert (1931)	?	Dutch	1.5-15	Range
Chiba (1935)	2	English	4.6	Average syllable rate; reading
	1	German	4.4	
	3	French	4.4	
	1	Russian	5.2	
	1	Korean	5.6	
	1 (?)	Japanese	7.7	
	1	Chinese	4.2	
	1	Hindustani	5.3	
	1	Mongolian	4.2	

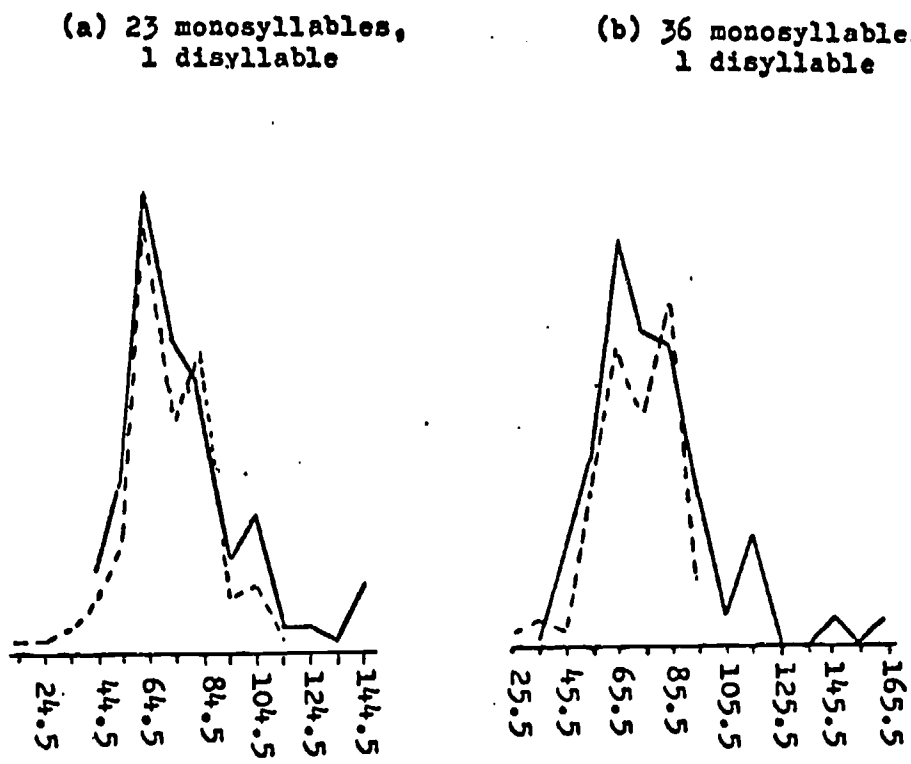


Figure 54. Average syllable duration frequency histograms from different speaking conditions. Solid line: 116 men; dashed line: 100 women. Abscissa: average durations of syllables, in centiseconds. Ordinate: frequency of occurrence (uncalibrated). Histogram for men's values in Figure 54h corrected from original. (After Kaiser 1939)

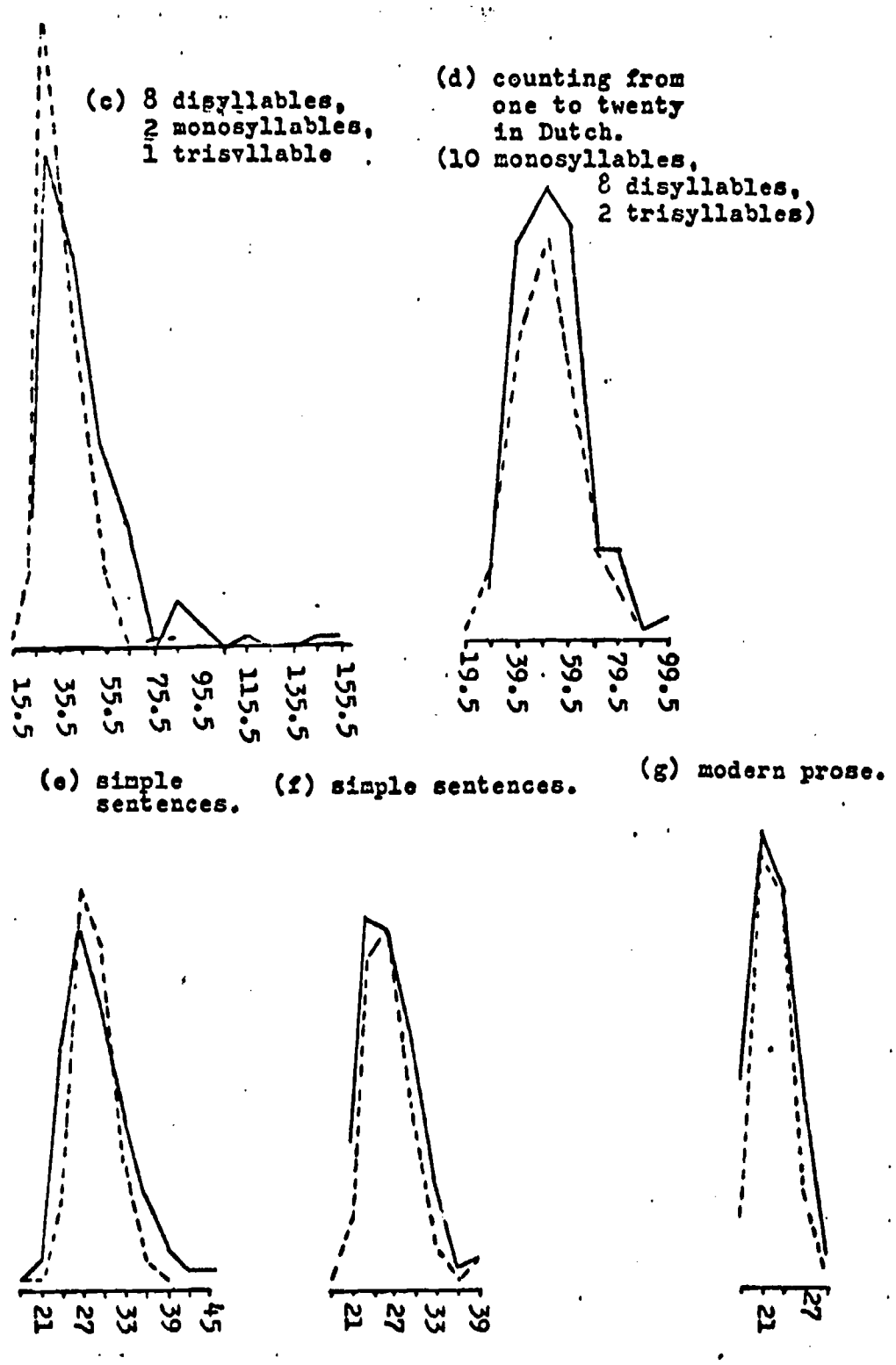


Figure 54 (continued).

Spontaneous Speech

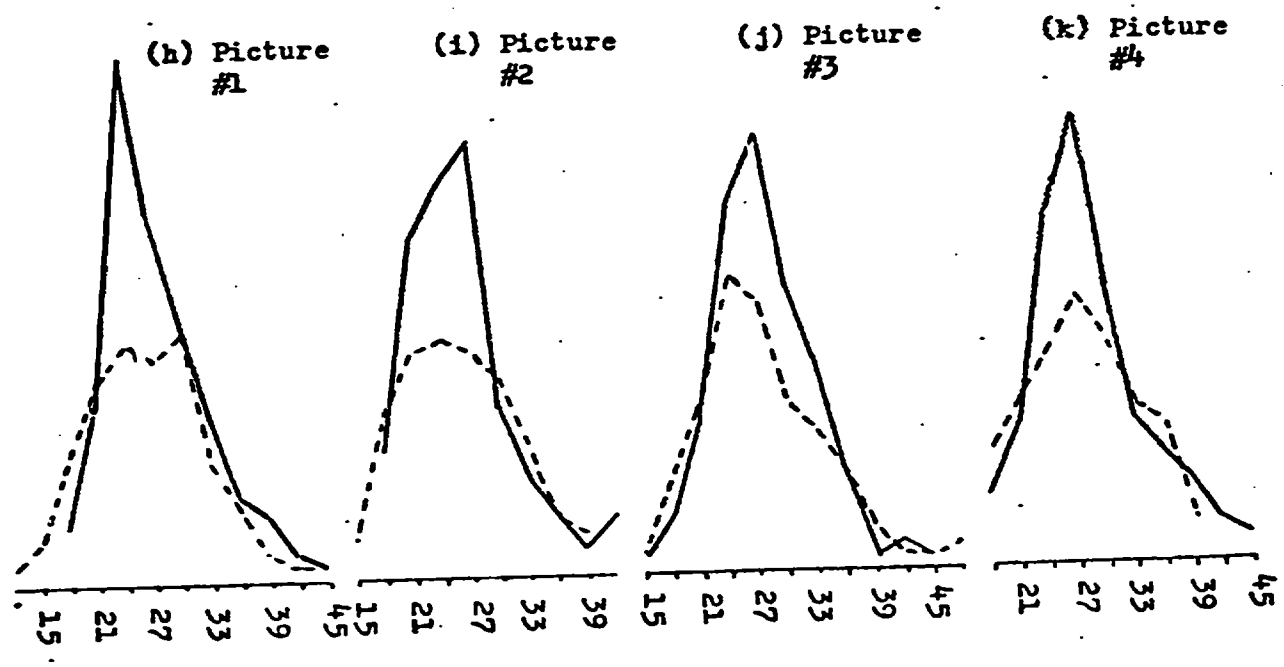


Figure 54 (continued).

determined by counting the number of syllables in a given passage and then dividing by the duration of the passage (which was determined by a stopwatch), but since many pauses were not subtracted from the passages before the calculation of average syllable duration was made the values for other than continuous speech are somewhat suspect, as was noted by Kaiser herself. Still the data is of interest. Figure 54 gives some of the *average syllable duration frequency histograms* obtained under the various speaking conditions for 116 male students and 100 female students. (That is, the average syllable duration for each of 216 students was determined and graphed as histograms with 116 counts for the males and 100 counts for the females.) Table XII gives the means and standard deviations for each of these histograms.

Table XII. Mean and Standard Deviation of Average Rate of Syllable Production for Different Speaking Conditions. (From Kaiser 1939; to accompany Figure 54) (Values in centiseconds)

	Men		Women	
Speaking Condition	Mean	S.D.	Mean	S.D.
(a)	78.2	22.5	71.4	17.1
(b)	74.4	17.8	70.7	14.1
(c)	47.4	----	38.5	----
(d)	43.0	13.6	40.7	12.9
(e)	29.3	5.0	28.7	3.2
(f)	27.0	4.3	26.6	3.6
(g)	22.5	3.0	22.7	2.5
(h)	27.6	5.5	26.4	6.2
(i)	25.8	4.2	25.8	5.8
(j)	27.6	5.2	27.6	6.6
(k)	28.3	6.0	26.5	5.6

Since the average values decrease when more di-syllabic words are included, it suggests that the average duration of such units depends on how much there is to say in the immediate future of what is currently being spoken, which has been found to be the case by Gaitenby (1965) and Lindblom (1968b).

Looking at all this, there is no obvious evidence which would point to "the magical one-sixth of a second as a basic time unit in speech production." It is important for Lenneberg's hypothesis that the rate of syllables -- or some units of articulation -- be found to have an average rate of around 1/6 sec. On this hangs the usefulness of the evidence from both (a) the critical delay causing maximum disruption of speech in delayed auditory feedback and (b) that rate of switching the speech signal from ear to ear or the rate of interrupting speech, which maximally reduces the intelligibility of the signal (Cherry and Taylor 1954, Huggins 1964).

However, it is perhaps wrong to quibble about the cited evidence since Lenneberg proposes the kind of experimental evidence that is needed to support or reject his hypothesis (a commendable practice too seldom followed by others in the field):

> The statistic necessary to prove or reject our hypothesis is quite simple. At present the only obstacle is the necessity of making observations and measurements of hundreds of thousands of events. Suppose we programmed an electronic computer to search the electrical analogue of a speech signal for that point in time at which any voiceless stop is released, and then measured the time lapse between all such successive points. From these data we can make histograms (bar-charts) showing the frequency distribution of all measurements. Since our hypothesis assumes that the variable *syllable-duration-time* is not continuous and that there are time quanta, the frequency distribution should be multimodal; and since the basic time unit is predicted to be 160 $\pm$ 20 msec., the distance between the peaks should be equal to or multiples of this unit. [p. 119]

The reason for looking for multi-modal distribution of the interval durations is to allow for the possibility of missing some intervals and catching instead intervals which are really twice, three times or some other integral multiple of the basic period of 1/6 sec. It is for this reason that the histograms given by Kaiser in Figure 54 cannot be used for this purpose. They represent averages of syllable rate not individual syllable durations. The peaks that do occur, some of them, interestingly, at integral multiples of 1/6 sec., indicate only that certain subjects had an average rate of 4/6 sec., others at 5/6 sec., etc.

An experiment similar to the one Lenneberg asks for was attempted.* Although it still calls for some further refinement and has yet to be run with a sufficiently large population (the single subject was, again, JO), the results do represent a tallying of about 10,000 separate intervals. It shall be reported here in the hope that other researchers will find it sufficiently interesting to want to try to replicate it.

Scanning the acoustic waveform of speech in search of the moment of release of voiceless stops is not so easy, so the measurements (by a computer) were made instead on the intervals between successive maxima in jaw movements (transduced by the device described on page 138). From the point of view of the computer program this meant detecting zero-crossings, in a given direction, of the first derivative of the signal representing deflections of the jaw. The units of time counted

---

* This experiment was presented orally in Ohala et al. 1969.

were (1.7)/(512) sec. or 3.3 msec., however the accuracy was somewhat less than this because in order not to count intervals due to noise the signal was required to cross the zero line and a threshold line before the interval would be counted. Thus the estimated accuracy of the time measurement is ± 10 msec. Intervals exceeding 1.7 seconds were not counted. However, using this program there was no way to ignore pauses, so if they were less than 1.7 sec. they were included. It is doubtful that this seriously affected the results. The subject read aloud technical material at a rate comfortable to him for about a total of 1 1/2 hours with short resting periods every 10 minutes or so. The resulting interval histogram representing a little over 10,000 intervals is shown in Figure 55.

One might first note that the shape of the histogram is quite smooth, almost like a Poisson distribution. Counts in the range from 0 to just under 100 msec. undoubtedly represent noise: some unwanted mechanical vibration of the light attached to the subject's lower teeth. If there are multiple peaks here they do not stand out very well. One might possibly be able to make a case for there being two peaks, one around 220 msec. and one around 310 msec. (which, as it turns out, is in agreement with Kaiser's data on continuous speech). However they are not separated by 160 msec., nor is either one within the predicted range of 140-200 msec. Only the second "peak" is close to an integral multiple of 160 msec. Nor does this histogram provide any evidence for any other rhythm. The two "peaks" are separated by about 90 msec., but neither 220 nor 310 msec. are integral multiples of 90, which would be expected if the basic period of the rhythm were 90 msec., i.e., a rhythm of about 11 c.p.s. This is far from being conclusive, however. It is necessary to try this out using a device that can detect "events" (defined somehow) that are produced by more than one articulator. That way shorter intervals can be obtained and the 160 msec. rhythm, if it exists, may show up better. Also it should be tried with spontaneous speech which may be different from "reading aloud" in that the subject generates his own rhythms, not those dictated by previously written material.

If one sees only one peak in this interval histogram it would be around 250-300 msec. and this may represent either the favored syllable duration or the favored rate at which the jaw moves. Very likely the tongue or lips would have a much smaller favored interval, and the respiratory muscles a much larger favored interval. Clearly more research is needed. For now there is no positive basis for accepting or rejecting Lenneberg's 1/6 sec. hypothesis, however, the evidence is not wholly favorable to it.

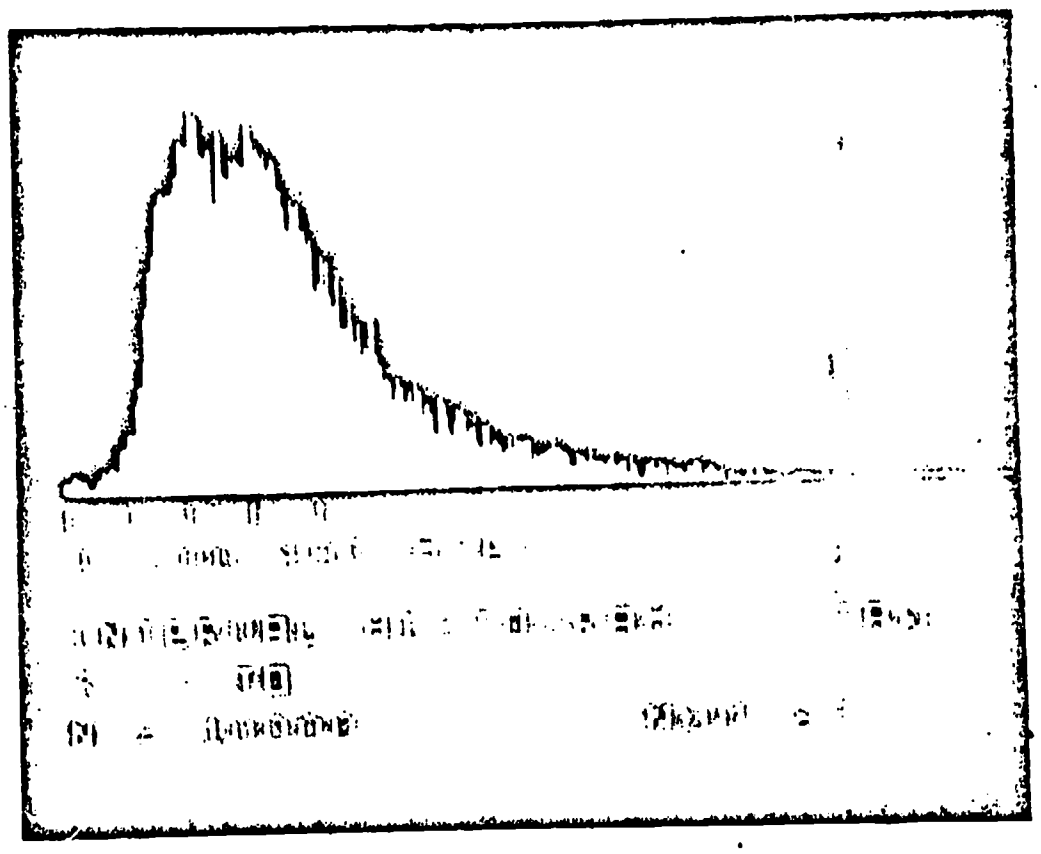


Figure 55.

Appendix: Illustrations of the Anatomy Relevant to the Discussion in Chapter 2

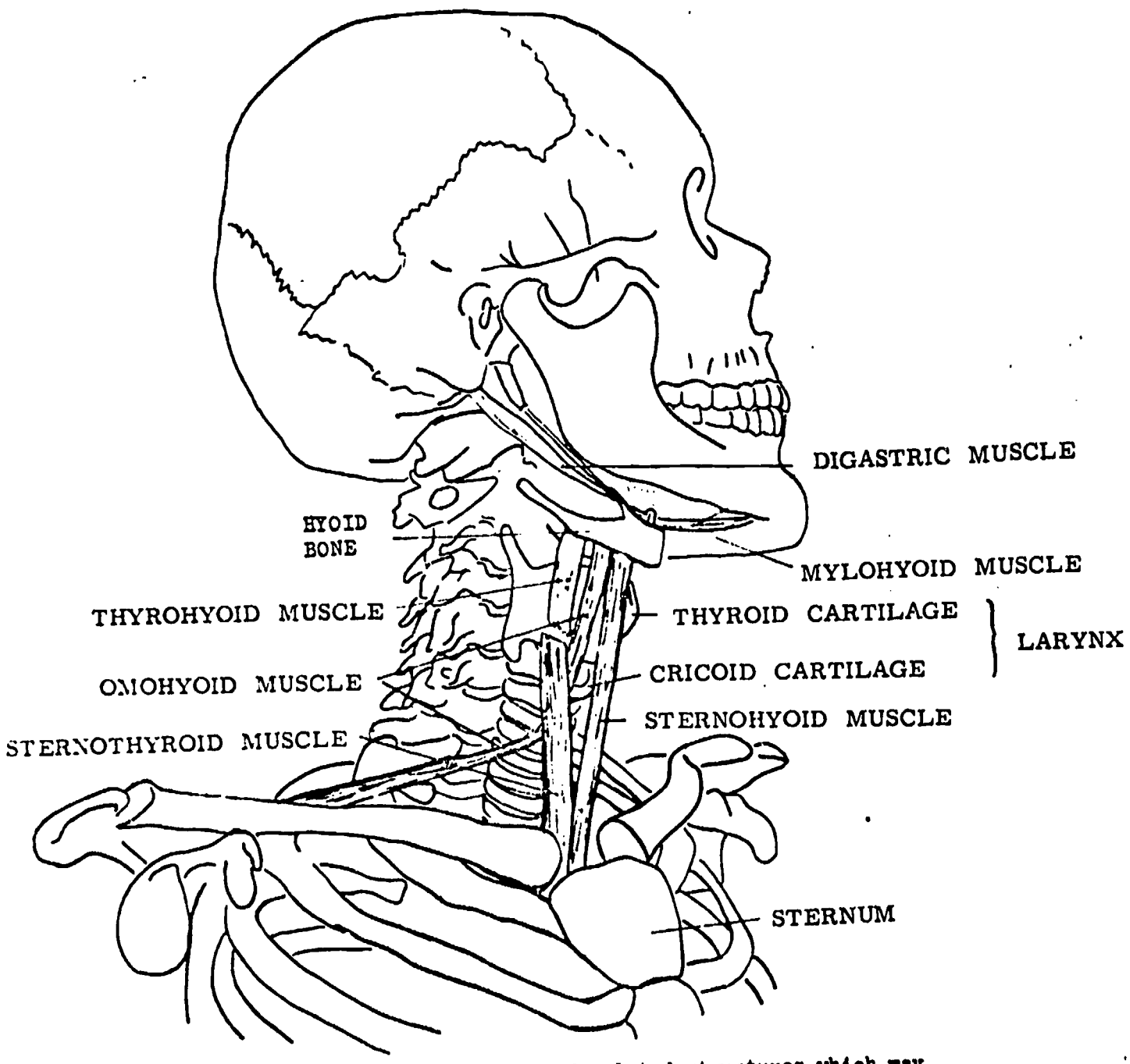


Figure 56. Sketch of the muscles and related structures which may contribute to the gross movements of the larynx. Muscles on the left side are not shown but are symmetrical to those on the right. As can be seen, the sternohyoid muscle is not attached directly to the larynx and can lower the larynx only indirectly by lowering the hyoid bone which is attached to the larynx through membranous and ligamentous connections. Structurally it would seem that the thyrohyoid and sternothyroid muscles would be more directly involved in raising or lowering the larynx, respectively. (Skeletel frame traced from Pernkopf)

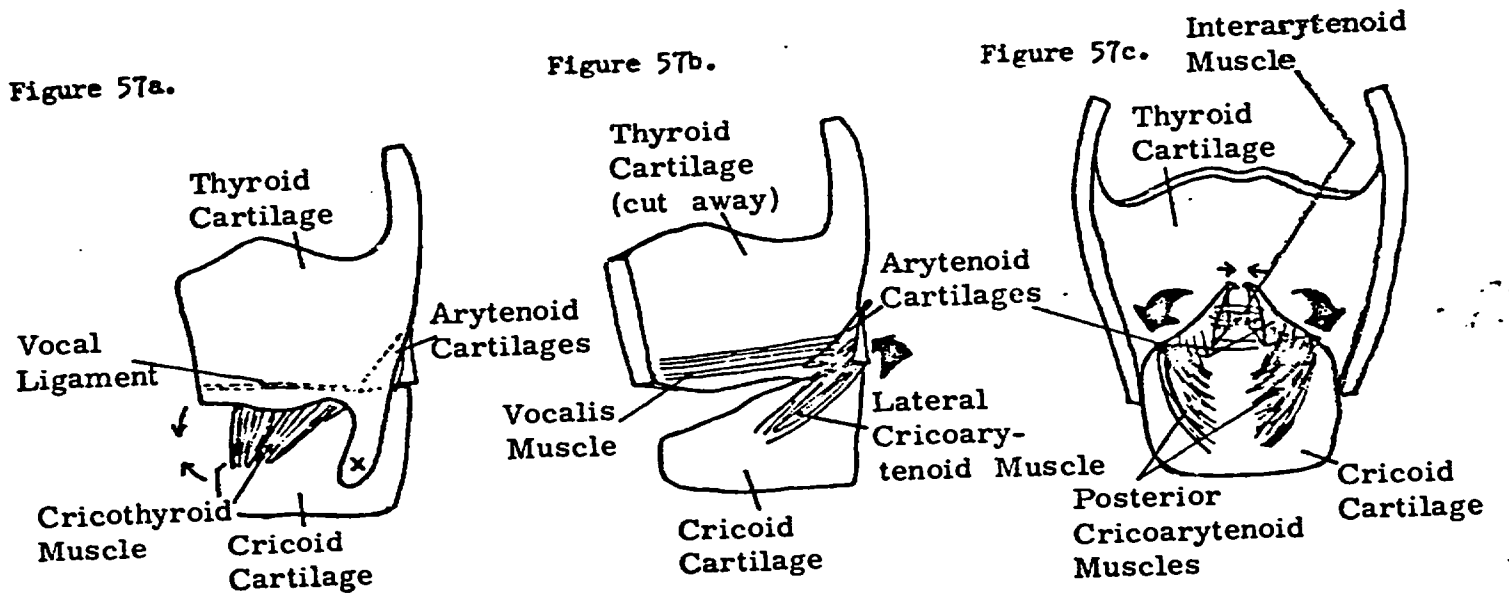


Figure 57. Schematic sketches of the larynx and its muscles.

Figure 57a. Left side view of the larynx showing the cricothyroid muscle. Upon contraction this muscle draws the front portions of the thyroid and cricoid cartilages closer together (arrows) as the thyroid cartilage pivots on the cricoid cartilage at the spot marked with an "x." This causes a stretching and increase in tension in the vocal ligament (the edge of the vocal cords extending from the arytenoid cartilages to the inside of the front of the thyroid cartilage). During phonation this causes an increase in pitch.

Figure 57b. Left side view of the larynx with the left half of the thyroid cartilage cut away. The vocalis muscles attach to the inside of the thyroid cartilage and the front portions of the arytenoid cartilages. The lateral cricothyroid muscles attach to the sides of the arytenoid and cricoid cartilages; upon contraction the arytenoids rotate obliquely towards the midline (large arrow) bringing the vocal cords together for phonation.

Figure 57c. Rear view of the larynx. The interarytenoid muscles pull the arytenoid cartilages closer together (small arrows) and thus make the vocal cords more tightly adducted. The posterior cricoarytenoid muscles pull the arytenoids in the opposite direction (large arrows) to that accomplished by the lateral cricoarytenoid muscles. This abducts the vocal cords.

BIBLIOGRAPHY

ABERCROMBIE, D. 1967
ELEMENTS OF GENERAL PHONETICS
CHICAGO: ALDINE

ALLEN, G. D. 1968
THE PLACE OF RHYTHM IN A THEORY OF LANGUAGE
WORKING PAPERS IN PHONETICS [UCLA] 10:60-84

AMENOMORI, Y. 1960
SIGNIFICANCE OF THE HYOID BONE IN PHONATION [ANALYSIS OF THE SUSPENSION MECHANISMS OF THE LARYNX]. [IN JAPANESE]
OTO-RHINO-LARYNGOLOGY CLINIC [KYOTO] 53:117-44

AMENOMORI, Y. 1961
SOME STUDIES ON THE 'SUSPENDING MECHANISM OF THE LARYNX.' [IN JAPANESE]
STUDIA PHONOLOGICA 1:95-110

ANDERSEN, H. C. 1837
KEJSERENS NYE KLAEDER [THE EMPEROR'S NEW CLOTHES]. [IN DANISH]
IN: EVENTYR, FORTALTE FOR BOERN
COPENHAGEN: C. A. REITZEL

ANDREWS, R. J. 1965
THE ORIGINS OF FACIAL EXPRESSIONS
SCIENTIFIC AMERICAN 213.4:88-94

ANTHONY, J. K. F. 1968
STUDY OF THE LARYNX II
WORK IN PROGRESS [UNIVERSITY OF EDINBURGH] 2:77-82

ARDRAN, G. M., WULSTAN, D. 1967
THE ALTO OR COUNTERTENOR VOICE
MUSIC AND LETTERS 48:17-22

ARMSTRONG, L. E., WARD, I. C. 1926
A HANDBOOK OF ENGLISH INTONATION
CAMBRIDGE: HEFFER

ARNOLD, G. E. 1961
PHYSIOLOGY AND PATHOLOGY OF THE CRICOTHYROID MUSCLE
LARYNGOSCOPE 71:687-753

BASMAJIAN, J. V., STECKO, G. 1962
A NEW BIPOLAR ELECTRODE FOR ELECTROMYOGRAPHY
JOURNAL OF APPLIED PHYSIOLOGY 17:849

BERG, JW. VAN DEN 1956
DIRECT AND INDIRECT DETERMINATION OF THE MEAN SUBGLOTTIC PRESSURE
FOLIA PHONIATRICA 8:1-24

BERG, JW. VAN DEN 1957
SUBGLOTTIC PRESSURES AND VIBRATIONS OF THE VOCAL FOLDS
FOLIA PHONIATRICA 9:65-71

BERG, JW. VAN DEN, TAN, T. S. 1959
RESULTS OF EXPERIMENTS WITH HUMAN LARYNXES
PRACTICA OTO-RHINO-LARYNGOLOGICA 21:425-50

BERNSTEIN, N. A. 1935
THE PROBLEM OF THE INTERRELATION OF CO-ORDINATION AND LOCALIZATION
ARHIV BIOLOSHKIH NAUKA [BELGRADE] 38

BERNSTEIN, N. A. 1967
THE COORDINATION AND REGULATION OF MOVEMENTS
OXFORD: PERGAMON PRESS

BOLINGER, D. L. 1958
A THEORY OF PITCH ACCENT IN ENGLISH
WORD 14:109-49

BOLINGER, D. L. 1964
INTONATION AS A UNIVERSAL
IN: PROCEEDINGS OF THE NINTH INTERNATIONAL CONGRESS OF LINGUISTS, PP 833-44
EDITED BY: LUNT, H.
THE HAGUE: MOUTON

BOLINGER, D. L. 1965
FORMS OF ENGLISH: ACCENT, MORPHEME, ORDER
EDITED BY: ABE, I., KANEKIYO, T.
CAMBRIDGE: HARVARD UNIVERSITY PRESS

BOSMA, J. F. 1964
RESPIRATORY MOTION PATTERNS OF THE NEWBORN INFANT CRY
IN: PHYSICAL DIAGNOSIS OF THE NEWLY BORN, PP 103-16
EDITED BY: KAY, J. L.
COLUMBUS: ROSS LABORATORIES

BOSMA, J. F., TRUBY, H. M., LIND, J. 1965
CRY MOTIONS OF THE NEWBORN INFANT
IN: NEWBORN INFANT CRY, PP 61-92
EDITED BY: LIND, J.
UPPSALA: ALMQVIST AND WIKSELLS

BOUHUYS, A. EDITOR. 1968
SOUND PRODUCTION IN MAN
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1:1-381

BOUHUYS, A., PROCTOR, D. F., MEAD, J. 1966
KINETIC ASPECTS OF SINGING
JOURNAL OF APPLIED PHYSIOLOGY 21:483-96

BOYD, J. A. 1965
DIFFERENCES IN THE DIAMETER AND CONDUCTION VELOCITY OF MOTOR AND FUSIMOTOR FIBRES IN NERVES TO DIFFERENT MUSCLES IN THE HIND LIMB OF THE CAT
IN: STUDIES IN PHYSIOLOGY PRESENTED TO J. C. ECCLES, PP 7-12
EDITED BY: CURTIS, D. R., MCINTYRE, A. K.
NEW YORK: SPRINGER-VERLAG

BROAD, D. J. 1968
SOME PHYSIOLOGICAL PARAMETERS FOR PROSODIC DESCRIPTION
SANTA BARBARA: SPEECH COMMUNICATIONS RESEARCH LABORATORIES

CHERRY, E. C., TAYLOR, W. K. 1954
SOME FURTHER EXPERIMENTS UPON THE RECOGNITION OF SPEECH, WITH ONE AND WITH TWO EARS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 26:554-59

CHIBA, T. 1935
A STUDY OF ACCENT
TOKYO: FUZANBO

CHISTOVICH, LUDMILLA A. 1969
VARIATION OF THE FUNDAMENTAL VOICE PITCH AS A DISCRIMINATORY CUE FOR CONSONANTS
SOVIET PHYSICS-ACOUSTICS 14:372-78

CHISTOVICH, LUDMILLA A., KOZHEVNIKOV, V. A., ALYAKRINSKII, V. V., BONDARKO, L. V., GOLUZINA, A. G., KLAAS, YU. A., KUZMIN, YU. I., LISENKO, D. M., LUBLINSKAYA, V. V., FEDOROVA, N. A., SHUPLYAKOV, V. S., SHUPLYAKOVA, R. M. 1965
RECH: ARTIKULYATSIYA I VOSPRIYATIYE [SPEECH: ARTICULATION AND PERCEPTION]
MOSCOW: NAUKA [ENGLISH TRANSLATION AVAILABLE THROUGH THE U. S. DEPARTMENT OF COMMERCE, # JPRS 30,543]

CHOMSKY, N. 1964
CURRENT ISSUES IN LINGUISTIC THEORY
IN: THE STRUCTURE OF LANGUAGE, PP 50-116
EDITED BY: FODOR, J. A., KATZ, J. J.
ENGLEWOOD CLIFFS: PRENTICE-HALL

CHOMSKY, N. 1965
ASPECTS OF THE THEORY OF SYNTAX
CAMBRIDGE: MIT PRESS

CHOMSKY, N. 1967A
DISCUSSION IN: BRAIN MECHANISM UNDERLYING SPEECH AND LANGUAGE, P 100
EDITED BY: DARLEY, F. L.
NEW YORK: GRUNE AND STRATTON

CHOMSKY, N. 1967B
THE FORMAL NATURE OF LANGUAGE
IN: BIOLOGICAL FOUNDATIONS OF LANGUAGE, PP 397-442
BY: LENNEBERG, E.
NEW YORK: JOHN WILEY AND SONS

CHOMSKY, N., HALLE, M. 1968
THE SOUND PATTERN OF ENGLISH
NEW YORK: HARPER AND ROW

CHOMSKY, N., HALLE, M., LUKOFF, F. 1956
ON ACCENT AND JUNCTURE IN ENGLISH
IN: FOR ROMAN JAKOBSON, PP 65-80
EDITED BY: HALLE, M., LUNT, H., MCCLEAN, H.
THE HAGUE: MOUTON

COLEMAN, H. O. 1914
INTONATION AND EMPHASIS
MISCELLANEA PHONETICA 1:6-26

COOPER, F. S. 1967
DISCUSSION IN: COMMUNICATING BY LANGUAGE: THE SPEECH PROCESS
EDITED BY: HOUSE, A. S.
BETHESDA: DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

CRITCHLEY, M., KUBIK, C. S. 1925
THE MECHANISM OF SPEECH AND DEGLUTITION IN PROGRESSIVE BULBAR PALSY
BRAIN 48:492-534

DANILOFF, R. G., AMERMAN, J., MOLL, K. L. 1968
AN INVESTIGATION OF THE TIMING AND SYNERGY OF JAW MOVEMENT AND LIP RETRACTION IN NORMAL SPEECH
PAPER PRESENTED AT THE FALL 1968 MEETING OF THE AMERICAN SPEECH AND HEARING ASSOCIATION

DARWIN, C. 1872
THE EXPRESSIONS OF EMOTION IN MAN AND ANIMALS
LONDON: J. MURRAY

DELATTRE, P. 1962
SOME FACTORS OF VOWEL DURATION AND THEIR CROSSLINGUISTIC VALIDITY
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 34:1141-43

DENES, P. 1955
EFFECT OF DURATION ON THE PERCEPTION OF VOICING
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 27:761-64

DRAPER, M. H., LADEFOGED, P., WHITTERIDGE, D. 1960
EXPIRATORY MUSCLES AND AIRFLOW DURING SPEECH
BRITISH MEDICAL JOURNAL 18 JUNE:1837-43

FAABORG-ANDERSEN, K. 1957
ELECTROMYOGRAPHIC INVESTIGATION OF INTRINSIC LARYNGEAL MUSCLES IN HUMANS
ACTA PHYSIOLOGICA SCANDINAVICA 41, SUPPLEMENT 140

FAABORG-ANDERSEN, K. 1965
ELECTROMYOGRAPHY OF LARYNGEAL MUSCLES IN HUMANS. TECHNICS AND RESULTS
AKTUELLE PROBLEME PHONIATRIE UND LOGOPEDIE, SUPPLEMENTUM FOLIA PHONIATRICA 3:1-72

FAABORG-ANDERSEN, K., SONNINEN, A. 1960
THE FUNCTION OF THE EXTRINSIC LARYNGEAL MUSCLES AT DIFFERENT PITCH
ACTA OTO-LARYNGOLOGICA 51.1-2:89-93

FAABORG-ANDERSEN, K., VENNARD, W. 1964
ELECTROMYOGRAPHY OF EXTRINSIC LARYNGEAL MUSCLES DURING PHONATION OF DIFFERENT VOWELS
ANNALS OF OTOLOGY, RHINOLOGY AND LARYNGOLOGY 73:248-54

FAIRBANKS, G. 1954
A THEORY OF THE SPEECH MECHANISM AS A SERVOSYSTEM
JOURNAL OF SPEECH AND HEARING DISORDERS 19:133-39

FAIRBANKS, G. 1955
SELECTIVE VOCAL EFFECTS OF DELAYED AUDITORY FEEDBACK
JOURNAL OF SPEECH AND HEARING DISORDERS 20:333-45

FANT, G. 1958
ACOUSTIC THEORY OF SPEECH PRODUCTION
STOCKHOLM: ROYAL INSTITUTE OF TECHNOLOGY

FARNSWORTH, D. W. 1940
HIGH SPEED MOTION PICTURES OF HUMAN VOCAL CORDS
BELL LABORATORIES RECORD 18:203-08

FERREIN, A. 1741
DE LA FORMATION DE LA VOIX DE L HOMME
MEMOIRES D ACADEMIE ROYALE DES SCIENCES 51:409-42

FINK, B. R. 1962
TENSOR MECHANISM OF THE VOCAL FOLDS
ANNALS OF OTOLOGY, RHINOLOGY AND LARYNGOLOGY 71:591-600

FINTOFT, K. 1961
THE DURATION OF SOME NORWEGIAN SPEECH SOUNDS
PHONETICA 7:19-39

FISCHER-JORGENSEN, ELI 1964
SOUND DURATION AND PLACE OF ARTICULATION
ZEITSCHRIFT FUER PHONETIK, SPRACHWISSENSCHAFT UND
KOMMUNIKATIONSFORSCHUNG 21:157-62

FISCHER-JORGENSEN, ELI, HANSEN, T. 1959
AN ELECTRICAL MANOMETER AND ITS USE IN PHONETIC RESEARCH
PHONETICA 4:43-53

FLANAGAN, J. L. 1968
STUDIES OF A VOCAL-CORD MODEL USING AN INTERACTIVE LABORATORY COMPUTER
IN: PREPRINTS OF THE SPEECH SYMPOSIUM, KYOTO, 1968

FLANAGAN, J. L., LANDGRAF, LORINDA 1967
SELF-OSCILLATING SOURCE FOR VOCAL-TRACT SYNTHESIZERS
IN: PREPRINTS OF THE 1967 CONFERENCE ON SPEECH COMMUNICATION AND PROCESSING (MIT)
REPRINTED IN: IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS AU-16:57-64 (1968)

FLANAGAN, J. L., LANDGRAF, LORINDA 1969
EXCITATION OF VOCAL-TRACT SYNTHESIZERS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 45:764-69

FOURNIE 1887
ESSAI DE PSYCHOLOGIE
PARIS

FRENCH, T. R. 1884
ON A PERFECTED METHOD OF PHOTOGRAPHING THE LARYNX
NEW YORK MEDICAL JOURNAL 40:653

FROMKIN, VICTORIA A. 1965
SOME PHONETIC SPECIFICATIONS OF LINGUISTIC UNITS: AN ELECTROMYOGRAPHIC INVESTIGATION
WORKING PAPERS IN PHONETICS [UCLA] 3

FROMKIN, VICTORIA A. 1968
SPECULATIONS ON PERFORMANCE MODELS
JOURNAL OF LINGUISTICS 4:47-68

FROMKIN, VICTORIA A., LADEFOGED, P. 1966
ELECTROMYOGRAPHY IN SPEECH RESEARCH
PHONETICA 15:219-42

FROMKIN, VICTORIA A., OHALA, J. 1968
LARYNGEAL CONTROL AND A MODEL OF SPEECH PRODUCTION
PREPRINTS OF THE SPEECH SYMPOSIUM, KYOTO, 1968

FRY, D. B. 1955
DURATION AND INTENSITY AS PHYSICAL CORRELATES OF LINGUISTIC STRESS
JOURNAL OF THE ACOUSTIC SOCIETY OF AMERICA 27:765FF

FRY, D. B. 1960
LINGUISTIC THEORY AND EXPERIMENTAL RESEARCH
TRANSACTIONS OF THE PHILOLOGICAL SOCIETY [LONDON] 1960:13-39

FRY, D. B. 1964
THE FUNCTION OF THE SYLLABLE
ZEITSCHRIFT FUER PHONETIK, SPRACHWISSENSCHAFT UND KOMMUNIKATIONSFORSCHUNG 17:215-21

FURUKAWA, M. 1967
A STUDY OF THE MECHANISM OF PHONATION USING EXISED LARYNGES [IN JAPANESE]
OTO-RHINO-LARYNGOLOGY CLINIC [KYOTO] 60:145-81

GAITENBY, JANE 1965
THE ELASTIC WORD
STATUS REPORT ON SPEECH RESEARCH [HASKINS LABORATORIES] 2--1965:3.1-.12

GARCIA, M. 1840
ECOLE DE GARCIA. TRAITE COMPLET DE L ART DU CHANT
PARIS: TROUPENAS

GRANDGENT, C. H. 1890
VOWEL MOVEMENTS
PUBLICATIONS OF THE MODERN LANGUAGE ASSOCIATION 5:148-74

GRIESMAN, B. L. 1943
MECHANISM OF PHONATION DEMONSTRATED BY PLANIGRAPHY OF THE LARYNX
ARCHIVES OF OTOLARYNGOLOGY 38:17-26

HADDING-KOCH, KERSTIN 1961
ACOUSTICO-PHONETIC STUDIES IN THE INTONATION OF SOUTHERN SWEDISH
LUND: GLEERUPS

HADDING-KOCH, KERSTIN, STUDDERT-KENNEDY, M. 1964
AN EXPERIMENTAL STUDY OF SOME INTONATION CONTOURS
PHONETICA 11:175-85

HAGGARD, M. P. 1969
PITCH IS VOICING--MANNER IS PLACE
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 46:97A

HALLE, M. 1962
PHONOLOGY IN GENERATIVE GRAMMAR
WORD 18:54-72

HALLE, M. 1964
ON THE BASES OF PHONOLOGY
IN: THE STRUCTURE OF LANGUAGE, PP 324-33
EDITED BY: FODOR, J. A., KATZ, J. J.
ENGLEWOOD CLIFFS: PRENTICE-HALL

HARSHMAN, R., LADEFOGED, P. 1967
THE LINC-8 COMPUTER IN SPEECH RESEARCH
WORKING PAPERS IN PHONETICS [UCLA] 7:57-68

HAST, M. H. 1961
SUBGLOTTIC AIR PRESSURE AND NEURAL STIMULATION IN PHONATION
JOURNAL OF APPLIED PHYSIOLOGY 16:1142-46

HELMHOLTZ, H. VON 1881
POPULAR LECTURES ON SCIENTIFIC SUBJECTS
LONDON: LONGMANS, GREEN

HENKE, W. 1967
PRELIMINARIES TO SPEECH SYNTHESIS BASED UPON AN ARTICULATORY MODEL
IN: PREPRINTS OF THE 1967 CONFERENCE ON SPEECH COMMUNICATION AND PROCESSING [MIT]

HENY, F. FORTHCOMING
DOCTORAL DISSERTATION [UCLA]

HILL, A. A. EDITOR. 1962A
FIRST TEXAS CONFERENCE ON PROBLEMS OF LINGUISTIC ANALYSIS IN ENGLISH
AUSTIN: UNIVERSITY OF TEXAS

HILL, A. A. EDITOR. 1962B
SECOND TEXAS CONFERENCE ON PROBLEMS OF LINGUISTIC ANALYSIS IN ENGLISH
AUSTIN: UNIVERSITY OF TEXAS

HIRANO, M., KOIKE, Y., VON LEDEN, H. 1967
THE STERNOHYOID MUSCLE DURING PHONATION
ACTA OTO-LARYNGOLOGICA 64:500-07

HIRANO, M., OHALA, J. 1969
USE OF HOOKED-WIRE ELECTRODES FOR ELECTROMYOGRAPHY OF THE INTRINSIC LARYNGEAL MUSCLES
JOURNAL OF SPEECH AND HEARING RESEARCH 12:362-73

HIRANO, M., OHALA, J., VENNARD, W. 1969
THE FUNCTION OF LARYNGEAL MUSCLES IN REGULATING FUNDAMENTAL FREQUENCY AND INTENSITY OF PHONATION
JOURNAL OF SPEECH AND HEARING RESEARCH 12:616-28

HIROTO, I., HIRANO, M., TOYOZUMI, Y., SHIN, T. 1962
A NEW METHOD OF PLACEMENT OF A NEEDLE ELECTRODE IN THE INTRINSIC LARYNGEAL MUSCLES FOR ELECTROMYOGRAPHY [IN JAPANESE]
OTO-RHINO-LARYNGOLOGY CLINIC [KYOTO] 55:499

HIROTO, I., HIRANO, M., TOYOZUMI, Y., SHIN, T. 1967
ELECTROMYOGRAPHIC INVESTIGATION OF THE INTRINSIC LARYNGEAL MUSCLES RELATED TO SPEECH SOUNDS
ANNALS OF OTOLOGY, RHINOLOGY AND LARYNGOLOGY 76:861-72

HOLLIEN, H., CURTIS, J. F. 1960
A LAMINAGRAPHIC STUDY OF VOCAL PITCH
JOURNAL OF SPEECH AND HEARING RESEARCH 3:361-71

HOLMES, J. N., MATTINGLY, I. G., SHEARME, J. N. 1964
SPEECH SYNTHESIS BY RULE
LANGUAGE AND SPEECH 7:127-43

HOSHIKO, M. S., BLOCKCOLSKY, VALEDA 1967
A RESPIROMETRIC STUDY OF LUNG FUNCTION DURING UTTERANCE OF VARYING SPEECH MATERIAL
SPEECH MONOGRAPHS 34:74-79

HOSOKAWA, H. 1961
PROPIOCEPTIVE INNERVATION OF STRIATED MUSCLES IN THE TERRITORY OF THE CRANIAL NERVES
TEXAS REPORTS ON BIOLOGY AND MEDICINE 19:405-64

HOUSE, A. S. EDITOR. 1967
COMMUNICATING BY LANGUAGE: THE SPEECH PROCESS
BETHESDA: DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

HOUSE, A. S., FAIRBANKS, G. 1953
THE INFLUENCE OF CONSONANT ENVIRONMENT UPON THE SECONDARY ACOUSTICAL CHARACTERISTICS OF VOWELS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 25:105-13

HUBLER, S. 1967
A HIGH INPUT IMPEDANCE ELECTROMYOGRAPHY PREAMPLIFIER
WORKING PAPERS IN PHONETICS [UCLA] 7:25-34

HUDGINS, C. V., STETSON, R. H. 1937
RELATIVE SPEED OF ARTICULATORY MOVEMENTS
ARCHIVES NEERLANDAISES DE PHONETIQUE EXPERIMENTALE 13:85-94

HUGGINS, A. W. F. 1964
DISTORTION OF THE TEMPORAL PATTERN OF SPEECH: INTERRUPTION AND ALTERNATION
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 36:1055-64

HUNT, C. C., PERL, E. R. 1960
SPINAL REFLEX MECHANISMS CONCERNED WITH SKELETAL MUSCLE
PHYSIOLOGICAL REVIEW 40:538-79

HURST, J. B. 1939
CONDUCTION VELOCITY AND DIAMETER OF NERVE FIBERS
AMERICAN JOURNAL OF PHYSIOLOGY 127:131-39

HUSSON, R. 1950
ETUDE DES PHENOMENES PHYSIOLOGIQUES ET ACOUSTIQUES FONDAMENTAUX DE LA VOIX CHANTEE
THESE DE LA FACULTE DE SCIENCE, PARIS
REVUE SCIENTIFIQUE 88:67-112, 131-46, 217-35

INOUE, O. 1931
TWO METHODS OF PITCH-RECORDING DEVICES [IN JAPANESE]
STUDY OF SOUNDS 4:57-60

ISSHIKI, N. 1959
REGULATORY MECHANISM OF THE PITCH AND VOLUME OF VOICE
OTO-RHINO-LARYNGOLOGY CLINIC [KYOTO] 52:1065-94

JONES, D. 1962
AN OUTLINE OF ENGLISH PHONETICS
CAMBRIDGE: HEFFER

KAISER, LOUISE 1939
BIOLOGICAL AND STATISTICAL RESEARCH CONCERNING THE SPEECH OF 216 DUTCH STUDENTS
ARCHIVES NEERLANDAISES DE PHONETIQUE EXPERIMENTALE 15:1-76

KATZ, J. J., POSTAL, P. M. 1964
AN INTEGRATED THEORY OF LINGUISTIC DESCRIPTIONS
CAMBRIDGE: MIT PRESS

KEENAN, J.S., BARRET, G. C. 1962
INTRALARYNGEAL RELATIONSHIPS DURING PITCH AND INTENSITY CHANGES
JOURNAL OF SPEECH AND HEARING RESEARCH 5:173-78

KELLY, J., GERSTMAN, L. 1961
AN ARTIFICIAL TALKER DRIVER FROM A PHONETIC INPUT
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 34:743[A]

KIM, C-W. 1966
THE LINGUISTIC SPECIFICATION OF SPEECH
WORKING PAPERS IN PHONETICS [UCLA] 5

KIMURA, N. 1961
ELECTROMYOGRAPHIC STUDY OF THE EXTRINSIC MUSCLE OF THE LARYNX
OTO-RHINO-LARYNGOLOGY CLINIC [KYOTO] 54:481

KOIKE, Y., PERKINS, W. H. 1968
APPLICATION OF A MINIATURIZED PRESSURE TRANSDUCER FOR EXPERIMENTAL SPEECH RESEARCH
FOLIA PHONIATRICA 20:360-68

KOYAMA, T., KAWASAKI, M., OGURA, J. H. 1969
MECHANICS OF VOICE PRODUCTION. I. REGULATION OF VOCAL INTENSITY
LARYNGOSCOPE 79:337-54

KOZHEVNIKOV, V. A., ARUTYUNYAN, E. A., BOROZDIN, A.N., VENTSOV, A. V., GRANSTREM, M. P., SHEIKIN, R. L., SHUPLYAKOV, V. C. 1966
METODY IZUCHENIYA RECHEVOGO DYKHANIYA
IN: MEKHANIZMY RECHEOBRAZOVANIYA I VOSPRIYATIYA SLOZHNYKH ZVUKOV, PP 45-55
EDITED BY: VENTSOV, A. V., KOZHEVNIKOV, V. A., KUZMIN, YU. I., CHISTOVICH, L. A.
MOSCOW: NAUKA

KRMPOTIC, JELENA 1959
DONNEES ANATOMIQUES ET HISTOLOGIQUES RELATIVES AUX
EFFECTEURS LARYNGO-PHARYNGO-BUCCAUX
REVUE DE LARYNGOLOGIE, OTOLOGIE, RHINOLOGIE 80:829-48

KUGELBERG, E. 1952
FACIAL REFLEXES
BRAIN 75:385-96

LADEFOGED, P. 1960
THE REGULATION OF SUB-GLOTTAL PRESSURE
FOLIA PHONIATRICA 12:169-75

LADEFOGED, P. 1962
SUBGLOTTAL ACTIVITY DURING SPEECH
IN: PROCEEDINGS OF THE FOURTH INTERNATIONAL CONGRESS OF
PHONETIC SCIENCES, PP 73-91
EDITED BY: SOVIJARVI, A., AALTO, P.
THE HAGUE: MOUTON

LADEFOGED, P. 1963
SOME PHYSIOLOGICAL PARAMETERS IN SPEECH
LANGUAGE AND SPEECH 6:109-19

LADEFOGED, P. 1964
A PHONETIC STUDY OF WEST AFRICAN LANGUAGES
CAMBRIDGE: CAMBRIDGE UNIVERSITY PRESS

LADEFOGED, P. 1967A
THREE AREAS OF EXPERIMENTAL PHONETICS
LONDON: OXFORD UNIVERSITY PRESS

LADEFOGED, P. 1967B
LINGUISTIC PHONETICS
WORKING PAPERS IN PHONETICS [UCLA] 6

LADEFOGED, P. 1968
LINGUISTIC ASPECTS OF RESPIRATORY PHENOMENA
IN: SOUND PRODUCTION IN MAN, PP 141-51
EDITED BY: BOUHUYS, A.
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1

LADEFOGED, P., DRAPER, M. H., WHITTERIDGE, D. 1958
SYLLABLES AND STRESS
MISCELLANEA PHONETICA 3:1-14

LADEFOGED, P., FROMKIN, VICTORIA A. 1968
EXPEPIMENTS ON COMPETENCE AND PERFORMANCE
IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS AU-16.1:130-
36

LADEFOGED, P., MCKINNEY, N. 1963
LOUDNESS, SOUND PRESSURE AND SUB-GLOTTAL PRESSURE IN SPEECH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 35:453-60

LANE, H. L. 1965
MOTOR THEORY OF SPEECH PERCEPTION: A CRITICAL REVIEW
PSYCHOLOGICAL REVIEW 72:275-310

LANE, H. L. 1967
A BEHAVIORAL BASIS FOR THE POLARITY PRINCIPLE IN LINGUISTICS
LANGUAGE 43:494-511

LANGUAGE LEARNING 1958
PROCEEDINGS OF THE CONFERENCE: LINGUISTICS AND THE TEACHING
OF ENGLISH AS A FOREIGN LANGUAGE, JULY 1957, ANN ARBOR,
MICHIGAN
LANGUAGE LEARNING, JUNE 1958

LASHLEY, K. S. 1951
THE PROBLEM OF SERIAL ORDER IN BEHAVIOR
IN: CEREBRAL MECHANISMS IN BEHAVIOR, PP 112-36
EDITED BY: JEFFRESS, L. A.
NEW YORK: WILEY

LAVER, J. 1968
PHONETICS AND THE BRAIN
WORK IN PROGRESS [UNIVERSITY OF EDINBURGH] 2:63-75

LAWRENCE, W. 1953
THE SYNTHESIS OF SPEECH FROM SIGNALS WHICH HAVE A LOW
INFORMATION RATE
IN: COMMUNICATION THEORY, PP 460-71
EDITED BY: JACKSON, W.
LONDON: BUTTERWORTHS

LEE, B. S. 1950
EFFECTS OF DELAYED SPEECH FEEDBACK
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 22:824-26

LEHISTE ILSE, PETERSON, G. E. 1961
SOME BASIC CONSIDERATIONS ON THE ANALYSIS OF INTONATION
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 33:419-25

LENNEBERG, E. H. 1967
BIOLOGICAL FOUNDATIONS OF LANGUAGE
NEW YORK: WILEY

LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S. 1958
THE ROLE OF SELECTED STIMULUS-VARIABLES IN THE PERCEPTION OF THE UNVOICED STOP CONSONANTS
AMERICAN JOURNAL OF PSYCHOLOGY 65:497-516

LIBERMAN, A. M., DELATTRE, P. C., GERSTMANN, L., COOPER, F. S. 1956
TEMPO OF FREQUENCY CHANGE AS A CUE FOR DISTINGUISHING CLASSES OF SPEECH SOUNDS
JOURNAL OF EXPERIMENTAL PSYCHOLOGY 52:127-37

LIEBERMAN, P. 1960
SOME ACOUSTICAL CORRELATES OF WORD STRESS IN AMERICAN ENGLISH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 32:451-54

LIEBERMAN, P. 1965
ON THE ACOUSTIC BASIS OF THE PERCEPTION OF INTONATION BY LINGUISTS
WORD 21:40-54

LIEBERMAN, P. 1966
SOME ACOUSTIC AND PHYSIOLOGIC CORRELATES OF THE BREATH GROUP
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 39:1218

LIEBERMAN, P. 1967A
INTONATION, PERCEPTION AND LANGUAGE
CAMBRIDGE: MIT PRESS

LIEBERMAN, P. 1967B
INTONATION AND THE SYNTACTIC PROCESSING OF SPEECH
IN: MODELS FOR THE PERCEPTION OF SPEECH AND VISUAL FORM, PP 314-19
EDITED BY: WATHEN-DUNN, W.
CAMBRIDGE: MIT PRESS

LIEBERMAN, P. 1968A
ON THE STRUCTURE OF PROSODY
ZEITSCHRIFT FUER PHONETIK, SPRACHWISSENSCHAFT UND KOMMUNIKATIONSFORSCHUNG 21:157-63

LIEBERMAN, P. 1968B
VOCAL CORD MOTION IN MAN
IN: SOUND PRODUCTION IN MAN, PP 28-38
EDITED BY: BOUHUYS, A.
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1

LIEBERMAN, P. 1968C
DIRECT COMPARISON OF SUBGLOTTAL AND ESOPHAGEAL PRESSURE DURING SPEECH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 43:1157-64

LIEBERMAN, P., GRIFFITHS, I. D., MEAD, J., KNUDSON, R.
1967
ABSENCE OF SYLLABIC CHEST PULSE
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 41:1614[A]

LIND, J. EDITOR. 1965
NEWBORN INFANT CRY
UPPSALA: ALMQVIST AND WIKSELLS

LINDBLOM, B. 1968A
VOWEL DURATION AND A MODEL OF LIP MANDIBLE COORDINATION
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM] 4--1967:1-29

LINDBLOM, B. 1968B
TEMPORAL ORGANIZATION OF SYLLABLE PRODUCTION
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM] 2-3--1968:1-15

LISKER, L. 1957
CLOSURE DURATION AND THE INTERVOCALIC VOICED-VOICELESS DISTINCTION IN ENGLISH
LANGUAGE 33:42-49

LISKER, L., ABRAMSON, A. S. 1964
A CROSS LANGUAGE STUDY OF VOICING IN INITIAL STOPS: ACOUSTICAL MEASUREMENTS
WORD 20:384-422

LISKER, L., ABRAMSON, A. S. 1967
SOME EFFECTS OF CONTEXT ON VOICE ONSET TIME IN ENGLISH STOPS
LANGUAGE AND SPEECH 10:1-28

LISKER, L., ABRAMSON, A. S. 1968
DISTINCTIVE FEATURES AND LARYNGEAL CONTROL
PAPER READ BEFORE THE 43RD ANNUAL MEETING OF THE LINGUISTIC SOCIETY OF AMERICA

LISKER, L., ABRAMSON, A. S., COOPER, F. S., SCHVEY, M. H.
1966
TRANSILLUMINATION OF THE LARYNX IN RUNNING SPEECH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 39:1218

LORENZ, K. Z. 1958
THE EVOLUTION OF BEHAVIOR
SCIENTIFIC AMERICAN 199.6:67-78

LUCAS KEENE, M. F. 1961
MUSCLE SPINDLES IN THE HUMAN LARYNGEAL MUSCLES
JOURNAL OF ANATOMY 95:25-29

LUCHSINGER, R., ARNOLD, G. E.
VOICE-SPEECH-LANGUAGE
BELMONT: WADSWORTH

MCCAWLEY, J. D. 1968
THE ROLE OF SEMANTICS IN A GRAMMAR
IN: UNIVERSALS IN LINGUISTIC THEORY, PP 124-69
EDITED BY: BACH, E., HARMS, R. T.
NEW YORK: HOLT, RINEHART AND WINSTON

MCCROSKEY, R. 1958
THE RELATIVE CONTRIBUTIONS OF AUDITORY AND TACTILE CUES TO CERTAIN ASPECTS OF SPEECH
SOUTHERN SPEECH JOURNAL 24:84-90

MACKAY, D. G. 1968
METAMORPHOSIS OF A CRITICAL INTERVAL: AGE-LINKED CHANGES IN THE DELAY OF AUDITORY FEEDBACK THAT PRODUCES MAXIMAL DISRUPTION OF SPEECH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 43:811-21

MCKINNEY, N. P. 1965
LARYNGEAL FREQUENCY ANALYSIS FOR LINGUISTIC RESEARCH
COMMUNICATION SCIENCES LABORATORY REPORT [ANN ARBOR] 14

MACNEILAGE, P. F. 1968
THE SERIAL ORDERING OF SPEECH SOUNDS
PROJECT ON LINGUISTIC ANALYSIS REPORTS [BERKELEY] SERIES 2, 8:M1-M52

MAJEWSKI, W., BLASDELL, R. 1969
INFLUENCE OF FUNDAMENTAL FREQUENCY CUES ON THE PERCEPTION OF SOME SYNTHETIC INTONATION CONTOURS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 45:450-57

MALECOT, A., PEEBLES, K. 1965
AN OPTICAL DEVICE FOR RECORDING GLOTTAL ADDUCTION-ABDUCTION DURING NORMAL SPEECH
ZEITSCHRIFT FUER PHONETIK, SPRACHWISSENSCHAFT UND KOMMUNIKATIONSFORSCHUNG 18:545-50

MANUKOVSKAIA, G. P. 1959
MODIFICATION OF THE MUSCLE INNERVATION PATTERN IN YOUNG ATHLETES IN THE PROCESS OF MASTERING GYMNASTIC EXERCISES
SECHENOV JOURNAL OF PHYSIOLOGY 45.11:28-33

MATHEWS, P. B. C. 1964
MUSCLE SPINDLES AND THEIR MOTOR CONTROL
PHYSIOLOGICAL REVIEWS 44:219-88

MATSUI, E., SUZUKI, T., UMEDA, NORIKO, OMURA, H. 1968
SYNTHESIS OF FAIRY TALES USING AN ANALOG VOCAL TRACT
REPORTS OF THE SIXTH INTERNATIONAL CONFERENCE ON ACOUSTICS
[TOKYO] B--159-62

MATTINGLY, I. G. 1966
SYNTHESIS BY RULE OF PROSODIC FEATURES
LANGUAGE AND SPEECH 9:1-13

MATTINGLY, I. G. 1968
SYNTHESIS BY RULE OF GENERAL AMERICAN ENGLISH
STATUS REPORT ON SPEECH RESEARCH [HASKINS LABORATORIES],
SUPPLEMENT, APRIL 1968

MEAD, J., BOUHUYS, A., PROCTOR, D. F. 1968
MECHANISMS GENERATING SUBGLOTTIC PRESSURE
IN: SOUND PRODUCTION IN MAN, PP 177-81
EDITED BY: BOUHUYS, A.
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1

MENZEL, P. FORTHCOMING
DOCTORAL DISSERTATION [UCLA]

MICHEL, R. 1954
DIE BEDEUTUNG DES MUSCULUS STERNOTHYREOIDEUS FUER DIE
RAHMENMODULATION DER MENSCHLICHEN STIMME
FOLIA PHONIATRICA 6:65-100

MINIFIE, F. D., KELSEY, C. A., HIXON, T. J. 1968
MEASUREMENTS OF VOCAL FOLD MOTION USING AN ULTRASONIC
DOPPLER VELOCITY MONITOR
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 43:1165-69

MITCHINSON, A. G. H., YOFFEY, J. M. 1948
CHANGES IN THE VOCAL FOLDS IN HUMMING LOW AND HIGH NOTES.
A RADIOGRAPHIC STUDY
JOURNAL OF ANATOMY 82:88-92

MOELLER, J., FISCHER, J. F. 1904
OBSERVATIONS ON THE ACTION OF THE CRICO-THYROIDEUS AND
THYRO-ARYTENOIDEUS INTERNUS
ANNALS OF OTOLOGY, RHINOLOGY AND LARYNGOLOGY 13:42-46

MOLL, K. L., DANILOFF, R. G. 1968
AN INVESTIGATION OF THE TIMING OF VELAR MOVEMENTS DURING
SPEECH
PAPER PRESENTED AT THE FALL 1968 MEETING OF THE AMERICAN
SPEECH AND HEARING ASSOCIATION

MUELLER, J. 1843
ELEMENTS OF PHYSIOLOGY
TRANSLATED BY: BALY, W.
ARRANGED BY: BELL, J.
PHILADELPHIA: LEA AND BLANCHARD

MUYSKENS, J. H. 1931
AN ANALYSIS OF ACCENT IN ENGLISH FROM KYMOGRAPH RECORDS
VOX 17.2:55-65

NORTHRUP, F. S. C. 1947
THE LOGIC OF THE SCIENCES AND HUMANITIES
NEW YORK: MACMILLAN

OERTEL 1895
DAS LARYNGO-STROBOSKOP UND DIE LARYNGOSTROBOSKOPISCHE UNTERSUCHUNG
ARCHIVE FUER LARYNGOLOGIE 3:1-16

OHALA, J. 1966
A NEW PHOTOELECTRIC GLOTTOGRAPH
WORKING PAPERS IN PHONETICS [UCLA] 4:40-52

OHALA, J. 1967
STUDIES OF VARIATIONS IN GLOTTAL APERTURE USING PHOTOELECTRIC GLOTTOGRAPHY
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 41:1613

OHALA, J., HIKI, S., HUBLER, S., HARSHMAN, R. 1968
PHOTOELECTRIC METHODS OF TRANSDUCING LIP AND JAW MOVEMENTS IN SPEECH
WORKING PAPERS IN PHONETICS [UCLA] 10:135-44

OHALA, J., HIKI, S., HUBLER, S., HARSHMAN, R. 1969
TRANSDUCING JAW AND LIP MOVEMENTS IN SPEECH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 45:1324

OHALA, J., HIRANO, M. 1967
CONTROL MECHANISMS FOR THE SEQUENCING OF NEUROMUSCULAR EVENTS IN SPEECH
PREPRINTS OF THE 1967 CONFERENCE ON SPEECH COMMUNICATION AND PROCESSING [MIT]

OHALA, J., HIRANO, M., VENNARD, W. 1968
AN ELECTROMYOGRAPHIC STUDY OF LARYNGEAL ACTIVITY IN SPEECH AND SINGING
REPORTS OF THE SIXTH INTERNATIONAL CONGRESS ON ACOUSTICS [TOKYO] B--5-8

OHALA, J., LADEFOGED, P. 1969
SUBGLOTTAL PRESSURE VARIATIONS AND GLOTTAL FREQUENCY
PAPER PRESENTED AT THE NOVEMBER 1969 MEETING OF THE
ACOUSTICAL SOCIETY OF AMERICA

OHALA, J., VANDERSLICE, R. 1965
PHOTOGRAPHY OF STATES OF THE GLOTTIS
WORKING PAPERS IN PHONETICS [UCLA] 2:58-59

OHMAN, S. 1967
NUMERICAL MODEL OF COARTICULATION
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 41:310-20

OHMAN, S. 1968
WORD AND SENTENCE INTONATION: A QUANTITATIVE MODEL
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF
TECHNOLOGY, STOCKHOLM] 2-3--1967:20-54

OHMAN, S., LINDQVIST, J. 1966A
ANALYSIS-BY-SYNTHESIS OF PROSODIC PITCH CONTOURS
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF
TECHNOLOGY, STOCKHOLM] 4--1965:1-6

OHMAN, S., LINDQVIST, J. 1966B
INSTRUMENTATION FOR SUBGLOTTAL AND SUPRAGLOTTAL AIR PRESSURE
MEASUREMENTS
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF
TECHNOLOGY, STOCKHOLM] 1--1966:9-10

OHMAN, S., MARTENSSON, A., LEANDERSSON, R., PERSSON, A.
1967
CRICO-THYROID AND VOCALIS MUSCLE ACTIVITY IN THE PRODUCTION
OF SWEDISH TONAL ACCENTS: A PILOT STUDY
QUARTERLY PROGRESS AND STATUS REPORT [ROYAL INSTITUTE OF
TECHNOLOGY, STOCKHOLM] 2-3--1967:55-57

OSTWALD, P. F. 1963
SOUNDMAKING: THE ACOUSTIC COMMUNICATION OF EMOTION
SPRINGFIELD: CHARLES THOMAS

PALMER, H. E. 1922
ENGLISH INTONATION
CAMBRIDGE: HEFFER

PALMER, H. E., BLANDFORTH, F. G. 1939
A GRAMMAR OF SPOKEN ENGLISH [SECOND EDITION]
CAMBRIDGE: HEFFER

PERKINS, W. H., YANAGAHIRA, N. 1968
PARAMETERS OF VOICE PRODUCTION: I. SOME MECHANISMS FOR THE
REGULATION OF PITCH
JOURNAL OF SPEECH AND HEARING RESEARCH 11:246-67

PERSON, R. S. 1958
ELECTROMYOGRAPHICAL STUDY OF COORDINATION OF THE ACTIVITY OF HUMAN ANTAGONIST MUSCLES IN THE PROCESS OF DEVELOPING MOTOR HABITS [RUSSIAN TEXT]
ZHURNAL VYSSHEI NERVNOI DEJATELNOSTI IMENI I. P. PAVLOVA 8:17-27

PETERSON, G., LEHISTE, ILSE 1960
DURATION OF SYLLABLE NUCLEI IN ENGLISH
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 32:693-703

PIKE, K. L. 1945
THE INTONATION OF AMERICAN ENGLISH
ANN ARBOR: UNIVERSITY OF MICHIGAN PRESS

POPPER, K. R. 1959
THE LOGIC OF SCIENTIFIC DISCOVERY
NEW YORK: HARPER AND ROW

POSTAL, P. 1968
ASPECTS OF PHONOLOGICAL THEORY
NEW YORK: HARPER AND ROW

PROCTOR, D. F. 1968
THE PHYSIOLOGIC BASIS OF VOICE TRAINING
IN: SOUND PRODUCTION IN MAN, PP 208-28
EDITED BY: BOUHUYS, A.
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1

RABINER, L. R., LEVITT, H., ROSENBERG, A. E. 1969
INVESTIGATION OF STRESS PATTERNS FOR SPEECH SYNTHESIS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 45:92-101

RETHI, L. 1896
EXPERIMENTELLE UNTERSUCHUNGEN UEBER DEN SCHWINGUNGSTYPUS UND DEN MECHANISMUS DER STIMMBAENDER BEI DER FALSETT-STIMME
SITZUNGBERICHTE DER KAISERLICHEN AKADEMIE DER WISSENSCHAFTEN [VIENNA], MATHEMATISCH-NATURWISSENSCHAFTLICHE KLASSE 105.3:197-212

RINGEL, R. L., STEER, M. D. 1963
SOME EFFECTS OF TACTILE AND AUDITORY ALTERATIONS ON SPEECH OUTPUT
JOURNAL OF SPEECH AND HEARING RESEARCH 6:369-78

ROHRER, F. 1916
DER ZUSAMMENHANG DES ATEMKRAEFTE UND IHRE ABHAENGIGKEIT VOM DEHNUNGSYNOSTAND DER ATMUNGSORGANE
PFLUEGERS ARCHIV FUER DIE GESAMTE PHYSIOLOGIE DES MENSCHEN UND DER TIERE 165:419-44

ROUSSELOT, L ABBE 1924
PRINCIPES DE PHONETIQUE EXPERIMENTALE [NOUVELLE EDITION]
TWO VOLUMES
PARIS: H. DIDIER

RUSHWORTH, G. 1966
SOME FUNCTIONAL PROPERTIES OF DEEP FACIAL AFFERENTS
IN: CONTROL AND INNERVATION OF SKELETAL MUSCLE, PP 125-33
EDITED BY: ANDREW, B. L.
EDINBURGH AND LONDON: LIVINGSTONE

SAPORTA, S. 1965
REVIEW OF: PSYCHOLOGY, STUDY OF A SCIENCE--STUDY II
EDITED BY: KOCH, S.
LANGUAGE 41:95-100

SATO, T. 1958
ON THE DIFFERENCES IN TIME STRUCTURE OF VOICED AND UNVOICED
STOP CONSONANTS [IN JAPANESE]
JOURNAL OF THE ACOUSTICAL SOCIETY OF JAPAN 14:117-22

SAWASHIMA, M., HIROSE, H. 1968
NEW LARYNGOSCOPIC TECHNIQUE BY USE OF FIBER OPTICS
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 43:168-69

SAWASHIMA, M., SATO, M., FUNASAKA, S., TOTSUKA, G. 1958
ELECTROMYOGRAPHIC STUDY OF THE HUMAN LARYNX AND ITS CLINICAL
APPLICATION [IN JAPANESE]
JOURNAL OF THE OTO-RHINO-LARYNGOLOGICAL SOCIETY OF JAPAN
61:1357-64

SCHANE, S. A. 1969
NATURAL PHONOLOGICAL RULES
PAPER READ AT THE 1969 UCLA HISTORICAL LINGUISTICS
CONFERENCE

SCHATZ, CAROL D. 1954
THE ROLE OF CONTEXT IN THE PERCEPTION OF STOPS
LANGUAGE 30:47-56

SCHEUER, J. L. 1964
FIBRE SIZE FREQUENCY DISTRIBUTION IN NORMAL HUMAN LARYNGEAL
NERVE
JOURNAL OF ANATOMY 98:99-104

SCOTT, N. C. 1939
AN EXPERIMENT ON STRESS PERCEPTION
LE MAITRE PHONETIQUE 1939:44-45

SEN, A. C. 1936
AN EXPERIMENTAL STUDY OF BENGALI OCCLUSIVES
IN: PROCEEDINGS OF THE SECOND INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, PP 184-93
CAMBRIDGE: UNIVERSITY PRESS

SHEN, YAO, PETERSON, G. G. 1962
ISOCHRONISM IN ENGLISH
STUDIES IN LINGUISTICS, OCCASIONAL PAPER 9

SLEDD, J. 1955
REVIEW OF: OUTLINE OF ENGLISH STRUCTURE
BY: TRAGER, G. L., SMITH, H. L., JR.
LANGUAGE 31:312-45

SMITH, S. 1944
THE DANISH STOD
COPENHAGEN: KAIFER

SOKOLOWSKY, R. R. 1943
EFFECT OF THE EXTRINSIC LARYNGEAL MUSCLES ON VOICE PRODUCTION
ARCHIVES OF OTOLARYNGOLOGY 38:355-64

SONNINEN, A. 1954
IS THE LENGTH OF THE VOCAL CORDS THE SAME AT ALL DIFFERENT LEVELS OF SINGING?
ACTA OTOLARYNGOLICA, SUPPLEMENTUM 118:219-31

SONNINEN, A. A. 1956
THE ROLE OF THE EXTERNAL LARYNGEAL MUSCLES IN LENGTH-ADJUSTMENT OF THE VOCAL CORDS IN SINGING
ACTA OTO-LARYNGOLOGICA, SUPPLEMENTUM 130

SONNINEN, A. 1968
THE EXTERNAL FRAME FUNCTION IN THE CONTROL OF PITCH IN THE HUMAN VOICE
IN: SOUND PRODUCTION IN MAN, PP 68-89
EDITED BY: BOUHUYS, A.
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155.1

SPUEHLER, H. F. 1962
DELAYED SIDE TONE AND AUDITORY FLUTTER
JOURNAL OF SPEECH AND HEARING RESEARCH 5:124-32

STETSON, R. H. 1928
MOTOR PHONETICS
AMSTERDAM: NORTH HOLLAND PUBLISHING CO.

STEVENS, K. N. 1968
ACOUSTIC CORRELATES OF PLACE OF ARTICULATION FOR STOP AND FRICATIVE CONSONANTS
QUARTERLY PROGRESS REPORT [MIT] 89:199-205

STEVENS, K. N., HALLE, M. 1967
REMARKS ON ANALYSIS BY SYNTHESIS AND DISTINCTIVE FEATURES
IN: MODELS FOR THE PERCEPTION OF SPEECH AND VISUAL FORM, PP 88-102
CAMBRIDGE: MIT PRESS

STRENGER, F. 1959
METHODS FOR DIRECT AND INDIRECT MEASUREMENT OF THE SUB-GLOTTIC AIR PRESSURE IN PHONATION
STUDIA LINGUISTICA [LUND] 13:98-112

SWEET, H. 1877
A HANDBOOK OF PHONETICS
OXFORD: CLARENDON PRESS

SWEET, H. 1911
PHONETICS
IN: ENCYCLOPEDIA BRITTANICA, ELEVENTH EDITION, 21:458-67

TRAGER, G. L., SMITH, H. L., JR. 1951
AN OUTLINE OF ENGLISH STRUCTURE
NORMAN: BATTENBY PRESS
ALSO: STUDIES IN LINGUISTICS, OCCASIONAL PAPER 3

TRUBY, H. M., LIND, J. 1965
CRY SOUNDS OF THE NEWBORN INFANT
IN: NEW BORN INFANT CRY, PP 7-60
EDITED BY: LIND, J.
UPPSALA: ALMQVIST AND WIKSELLS

UMEDA, NORIKO, TERANISHI, R. 1966
PHONEMIC FEATURE AND VOCAL FEATURE--SYNTHESIS OF SPEECH SOUNDS, USING AN ACOUSTIC MODEL OF VOCAL TRACT
JOURNAL OF THE ACOUSTIC SOCIETY OF JAPAN 22:195-203

VANDERSLICE, R. 1967
LARYNX VERSUS LUNGS: CRICOTHYROMETER DATA REFUTING SOME RECENT CLAIMS CONCERNING INTONATION AND ARCHETYPALITY
WORKING PAPERS IN PHONETICS [UCLA] 7:69-79

VANDERSLICE, R. 1968A
SYNTHETIC ELOCUTION: CONSIDERATIONS IN AUTOMATIC ORTHOGRAPHIC-TO-PHONETIC CONVERSION OF ENGLISH WITH SPECIAL REFERENCE TO PROSODIC FEATURES
WORKING PAPERS IN PHONETICS [UCLA] 8

VANDERSLICE, R. 1968B
THE PROSODIC COMPONENT: LACUNA IN TRANSFORMATIONAL THEORY
RAND CORPORATION PUBLICATION P-3874

VENTSOV, A. V., KOZHEVNIKOV, V. A., KUZMIN, YU. I.,
CHISTOVICH, LUDMILLA A. EDITORS. 1966
MEKHANIZMY RECHEOBRAZOVANIYA I VOSPRIYATIYA SLOZHNIKH ZVUKOV
MOSCOW: NAUKA

WANG, W. S-Y. 1962
STRESS IN ENGLISH
LANGUAGE LEARNING 12:69-75

WANG, W. S-Y. 1967
PHONOLOGICAL FEATURES OF TONE
INTERNATIONAL JOURNAL OF AMERICAN LINGUISTICS 33:93-105

WANG, W. S-Y. 1968
THE BASIS OF SPEECH
PROJECT ON LINGUISTIC ANALYSIS (BERKELEY), SECOND SERIES, 4

WENRICK, J. E. 1931
STIMULATION OF MUSCLES IN THE LARYNX TILTING THYROID
DOWNWARD ON THE CRICOID LOWERS PITCH
ARCHIVES NEERLANDAISES DE PHONETIQUE EXPERIMENTALE 6:92-100

WOLFF, P. H. 1966
THE NATURAL HISTORY OF CRYING AND OTHER VOCALIZATIONS IN
EARLY INFANCY
IN: DETERMINANTS OF INFANT BEHAVIOR IV
EDITED BY: FOSS, B. M.
LONDON: MENTHUEN

YANAGIHARA, N., VON LEDEN, H. 1966
THE CRICOTHYROID MUSCLE DURING PHONATION
ANNALS OF OTOLOGY, RHINOLOGY, AND LARYNGOLOGY 75:987-1006

ZENKER, W. 1964
DISCUSSION IN: RESEARCH POTENTIALS IN VOICE PHYSIOLOGY
EDITED BY: BREWER, D. W.
NEW YORK: UNIVERSITY PUBLISHERS

ZENKER, W., ZENKER, A. 1960
UEBER DIE REGELUNG DER STIMMLIPPENSPANNUNG DURCH VON AUSSEN
EINGREIFENDE MECHANISMEN
FOLIA PHONIATRICA 12:1-36

ZIMMER, K. E. 1968
PSYCHOLOGICAL CORRELATES OF SOME TURKISH MORPHEME STRUCTURE CONDITIONS
PROJECT ON LINGUISTIC ANALYSIS REPORTS [BERKELEY], SECOND SERIES, 8:1-25

ZWAARDEMAKER, H. 1915
LEERBOEK DER PHYSIOLOGIE VOLUME 2
HAARLEM: DE ERVEN BOHR

Flisberg, K. and Lindholm, T. (1970) Electrical stimulation of the human recurrent laryngeal nerve during thyroid operation. Acta Otolaryngologica 263:63-67.

Fromkin, Victoria A. and Ohala, J. (1968) Laryngeal Control and a Model of Speech Production. Preprints of the Speech Symposium, Kyoto, 1968. Reprinted with oral version in: Working Papers in Phonetics, UCLA, 10:98-110.

Hixon, T. J., Mead, J., and Klatt, D. H. (1970) Influence of forced transglottal pressure changes on vocal fundamental frequency. Paper read at 80th meeting of the Acoustical Society of America, Houstion, Texas, 4 November 1970.

Martin, S. (1956) Review of Manual of Phonology by C. Hockett. Language 56.701.

Negus, V. E. (1949) The comparative anatomy and physiology of the larynx. New York: Hafner Publishing Company.

Peterson, G. E. (1958) Some observations on speech. Quarterly Journal of Speech 44.402-412.