A Study of Prosodic Features*

Philip Lieberman
Haskins Laboratories, New Haven+

We will discuss a number of recent advances in the study of the prosodic elements of speech. We will deliberately neglect many recent studies that are based on the unaided senses of a trained observer. We will instead concentrate on studies that would have not been possible in the absence of current techniques for acoustic, physiologic, anatomical, and perceptual experimentation. We shall, moreover, deal with the "linguistic" aspects of intonation. Charles Darwin (1872) noted that cries convey the emotional state of the organism in both man and animals. Darwin was, of course, concerned with the attributes of communication that are common to both man and all other animals, i.e., the nonlinguistic aspects of the suprasegmental prosodic features. The linguistic analysis of the suprasegmental features becomes quite complex inasmuch as these features are also used for the nonlinguistic aspects of speech communication.

We are reserving the term "linguistic" for language-relevant aspects of the suprasegmental features, i.e., those aspects that serve to convey meaning through the medium of language. Linguistic systems differ quite fundamentally from nonlinguistic systems of communication in that individual cries, or phonetic elements, have no inherent meaning. They derive a meaning only after syntactic and semantic analysis. The sound [m] has no inherent "meaning" in a linguistic system. In English, the word man does have a definite meaning. The sound [m] as it is used in producing the word man has, however, no particular meaning. It forms part of a phonetic "coding" of a word which is the "object" that has the linguistic meaning. The sound [m] can also be used to code other words, e.g., am, mama, etc.

The sound [m] can also have a nonlinguistic function. A particular speaker may, for example, use this sound outside the language system to convey a particular emotional state. It might signify that he is happy. The sound [m] as the speaker used it for this nonlinguistic function would have a definite fixed meaning. No syntactic analysis would be necessary to derive its meaning. Note that its meaning would also be idiosyncratic. The listener might or might not know that this sound signified that the particular speaker were happy. The nonlinguistic "meaning" of [m] would not form part of the English language.

The sounds used in human speech thus serve for communication at many levels. We can easily differentiate at least five factors that are transmitted by means of speech:

+Also, University of Connecticut, Storrs.
1) The speech signal conveys acoustic cues that serve to identify the individual speaker. This aspect of speech communication is quite important. When telephone systems are degraded to the point where they do not transmit these acoustic cues the public begins to complain (Flanagan, 1965). Animal communication also makes use of cries to identify individual animals to their progeny, mates, friends, and associates. Birds, for example, employ such identification signals (Beer, 1969). The acoustic signal also serves to identify the species (Marler and Hamilton, 1966; Greenewalt, 1968) but this aspect of acoustic communication is not relevant to human speech at the present time since Homo sapiens is the only living species that can produce the range of sounds used in speech (Lieberman, 1968b; Lieberman and Crelin, 1971).

The cues that humans use to identify particular speakers involve both the segmental and the suprasegmental phonetic features. Individual speakers indeed may employ different syntactic "styles" that involve different base structures and different optional transformations\(^1\) to convey similar semantic information. It is clear, however, that the suprasegmental features are quite important in establishing the identity of a particular speaker.\(^2\) When the fundamental frequency of a speaker is transposed (by using a Vocoder apparatus) (Flanagan, 1965) it becomes quite difficult to identify the speaker. Transposing the fundamental frequency, of course, changes the perceived pitch of the speaker's voice.

2) The speech signal conveys the linguistic background of the speaker. Individual languages involve language-specific phonetic and syntactic elements. At the phonetic level, language-specific implementation rules (Lieberman, 1970) are involved as well as specific feature ensembles that may be drawn from the set of universal features (Jakobson et al., 1952; Chomsky and Halle, 1968). There are apparently language-specific elements that are manifested in intonation (Ladefoged, 1967; Lieberman, 1967).

3) The speech signal conveys the sex of the speaker. In many languages this occurs at the phonetic level. The fundamental frequency of the speech signal is usually lower for male speakers. This reflects the longer vocal cords that males usually have (Negus, 1949). The vocal cords usually increase in length in males at puberty as the thyroid cartilage grows larger. Adult females also have lower fundamental frequencies than juvenile females since their larynges also grow larger. The great and abrupt increase in the length of the vocal cords is, however, a secondary sexual dimorphism in males. Other acoustic differences also manifest the sex of speakers. Male speakers of English, for example, seem to use lower formant frequencies than do females.

\(^1\) We will operate within the framework of a generative grammar (Chomsky, 1957, 1968) that makes use of phonemic and phonetic features (Jakobson et al., 1952; Chomsky and Halle, 1968; Postal, 1968).

\(^2\) The fine structure of the fundamental frequency of phonation can even play a part in transmitting the state of health of the speaker. Measurements of the variations in fundamental periodicity, "pitch perturbations," have been used as a diagnostic tool for the early detection of cancer of the larynx as well as other laryngeal pathologies (Lieberman, 1963).
(Peterson and Barney, 1952). These differences may reflect the larger size of male vocal tracts.

4) The speech signal conveys the emotional state of the speaker. Much of the "meaning" of speech is communicated at this level. When we listen to a speaker we may be as aware of the emotional content of the signal, which conveys the speaker's attitude toward the situation, as of the "linguistic" content. In many situations the "linguistic" content of the message, i.e., the part of the message conveyed by the words, is quite secondary. Stereotyped greetings, like Good morning, probably serve as vehicles for emotional information. Stereotyped messages, like the elevator operator's Step to the rear of the car please, may primarily serve as carriers that transmit the emotional state of the speaker.

The "tone" of the speaker's voice may indicate whether he is annoyed at the passengers, whether he is happy, etc. Unfortunately, the information conveyed by the "tone" of the speaker's voice is somewhat ambiguous. The listener really does not know whether an "angry" tone means that the elevator operator is angry at the passengers or that his breakfast was not edible or that his back aches. Emotional information is rarely specific. There are further difficulties insofar as certain emotional nuances are themselves stereotyped. Thus in Newark, New Jersey, and San Francisco, California, different prosodic patterns may signify disdain. The listener must be aware of the speaker's background and current social convention. The "primary" emotional attributes like extreme pain may indeed be stable (Darwin, 1872), but many of the emotional and attitudinal nuances are probably dialect specific. These dialect-specific aspects are in a sense paralinguistic. They are to a certain degree arbitrary. They thus are linguistic in the sense that the relation between "meaning" associated with a sound and the sound is arbitrary. However, they are not like the "linguistic" aspects of the speech signal insofar as there is a direct relation between these signals and their meanings. There is no morphophonemic or syntactic level.

5) The "linguistic" content of the speech signal is naturally of paramount interest to linguists. Certain aspects of the prosodic features convey linguistic information. Like all other phonetic elements, these prosodic features have no meaning in themselves. We shall direct our attention to a review of the state of current research on some of these prosodic features. We will attempt to limit our discussion to the linguistic aspects of speech. This often is difficult when one deals with the prosodic features of intonation, accent, and prominence. Many analyses, e.g., Pike (1945), have attempted to treat what we have termed levels 4 and 5 as an entity. We will, however, attempt to differentiate these aspects of the speech signal, though we recognize that there will always be some uncertainty as to whether a particular prosodic feature is a dialect-specific level 4 or a linguistic level 5 event. Indeed the process by which language develops may, in part, consist of level 4 to level 5 transitions for particular phonetic features, particular words, or syntactic processes. Note that we are not stating that levels 1 to 4 are unimportant.

The Suprasegmental Prosodic Features: Acoustic Correlates

It is convenient and reasonable to consider two elements in connection with the suprasegmental prosodic features. One element is the suprasegmental
sentence or phrase intonation, the other element is the class of features like accent and prominence. Whereas the scope of intonation is generally an entire sentence or phrase, the scope of prominence or accent is usually a single syllable, though longer stretches of speech can also be accented or assigned prominence.

The results of many experiments that have involved not only the analysis of speech but speech synthesis in connection with Vocoder equipment have shown that the primary acoustic cue that signals the intonation of an utterance is the fundamental-frequency contour. That is, the manner in which the fundamental-frequency contour varies with respect to time largely determines the intonation of the utterance. Other factors undoubtedly also play a role in defining the perceived intonation of the utterance. The amplitude of the speech signal, for example, generally decreases toward the end of an intonation contour. There also are strong indications that the duration of a syllable is a function of its position within the intonation contour. However, the fundamental frequency at certain points in the intonation contour (Denes, 1959) appears to be the most important acoustic correlate of intonation. We will return to this point again. Many phonetic analyses of intonation have created the impression that the fundamental-frequency contour is the only acoustic correlate of intonation. These analyses assign linguistic significance to minute variations in fundamental frequency throughout the entire sentence. They, in effect, assume that the fundamental-frequency contour must be specified in minute detail throughout the entire intonation contour. Recent experimental evidence, which we will discuss, suggests that this is not the case.

We will begin by discussing one phonetic feature, the breath-group, that describes some of the linguistic functions of intonation (Lieberman, 1967, 1970; Lieberman et al., 1970). We will also discuss other constructs that have been introduced to describe the same linguistic phenomena that the breath-group describes, as well as constructs that are necessary to describe still other linguistic phenomena. Virtually all phonetic and acoustic studies agree that certain acoustic patterns occur in speech that specify intonation patterns. The disagreements and uncertainty are with regard to a) what acoustic parameters are important, b) how these acoustic parameters are generated and controlled by the human speech-production apparatus and c) how many distinct patterns there are and what their linguistic significance is. A description of the acoustic correlates of the breath-group is therefore in order at this point since it will be equivalent at the acoustic level to many of these alternate theoretical constructs. The acoustic description will also hopefully clarify some of the concepts that we shall discuss.

3The commercial application of Vocoder equipment, which would offer significant economies on high cost circuits like the Atlantic cable, has been delayed for over thirty years by the deficiencies that exist in "pitch extractors" (Flanagan, 1963).

4The study presented by Isacenko and Schadlich (1963) is typical of a class of studies that assign linguistic significance to minute (5Hz) variations in fundamental frequency over a long utterance.
In Figure 1 we have reproduced some data (Lieberman, 1967) that shows some of the acoustic parameters associated with a normal breath-group. (An equivalent notation is -breath-group.) The upper plot in this figure is a quantized sound spectrogram. The darkened areas that are enclosed by "contour lines" represent the relative energy that is present at the frequency plotted on the ordinate scale as a function of time. Time is plotted on the abscissa in seconds. The energy present in a timing pulse is displayed at the two points marked by arrowheads on the abscissa after 0.5 and 1.5 seconds. The speaker uttered the sentence, Joe ate his soup. Note, for example, the energy concentration at approximately 200 Hz (a Hz is equivalent to a cycle per second) at t = 0.4 seconds which is the spectrogram's representation of the first formant of the vowel of the word Joe. It is possible to determine relative energy levels by means of the quantized spectrogram.

The second plot from the top in Figure 1 is the smoothed fundamental frequency of phonation as a function of time. The fundamental frequency was derived by measuring the tenth harmonic on a narrow bandwidth spectrogram.

The uppermost plots in Figure 1 thus show that the fundamental frequency of phonation and sound energy both decrease at the end of the -breath-group. The duration of segmental phonemes also appears to increase at the end of the breath-group. (Note the duration of the closure interval of the stop /p/, which is longer than the closure interval of the stop /t/.) Also note that the relative energy balance changes at the end of the breath-group. (There is less energy in the higher formants of the vowel of soup as the breath-group ends.)

These observations will become more meaningful as we discuss the results of current research on the behavior of the larynx. We will also discuss the significance of the two lower plots in Figure 1. Similar acoustic correlates of what we have termed the normal breath-group have been reported by Chiba (1935) in an early study which made use of electronic analysis equipment.

Recent studies by Jassem (1959), Hadding-Koch (1961), Revtova (1965), Fromkin and Ohala (1968), Matsui et al. (1968), Ohala (1970), Vanderslice (1970), and Lieberman et al. (1970) also have derived similar acoustic correlates. Armstrong and Ward (1926) and Jones (1932) also postulate similar acoustic correlates for Tune I. The Jones and Armstrong and Ward studies, of course, were conducted without the benefit of modern instrumentation.

Some of the acoustic correlates of the normal breath-group may also be seen in Jones (1909) and Cowan (1936) where perceived pitch and fundamental frequency, respectively, are plotted as functions of time for relatively large

---

5The wide-band filter of the sound spectrographs that are usually used for the analysis of speech has a bandwidth of 300 Hz. This bandwidth accepts at least two harmonics of the glottal source for typical male speakers. The wide-band filter thus will manifest the formant frequencies rather than the individual harmonics (Flanagan, 1965). The formant frequencies are uniquely determined by the cross-sectional area function of the supralaryngeal vocal tract.
Acoustic and physiologic data for a normal, _breath_-group (after Lieberman, 1967)
data samples. Pike (1945) also postulates similar acoustic correlates for the "final pause [\]." The Trager and Smith (1951) "terminal juncture [#]" also appears to have similar acoustic correlates (Hadding-Koch, 1961; Lieberman, 1965).

Note that the acoustic correlates that we have discussed all are relative measurements over a span of time that encompasses a string of segmental phonetic elements. The normal breath-group is thus a true suprasegmental. The "pause" notation developed by Pike (1945) and Trager and Smith (1951) also is an implicit suprasegmental since the amplitude and fundamental frequency of the pause are relative measures that are defined with respect to the pitch-amplitude contour that precedes the terminal. The contour that precedes the terminal thus is an intimate part of the terminal contour. The Jones (1932) and Armstrong and Ward (1926) Tune I notation, of course, is an explicit suprasegmental phonetic entity. Harris (1944) also postulates, without any experimental evidence, suprasegmental intonation "morphemes."

In Figure 2 we have reproduced some data that shows some of the acoustic parameters associated with a +breath-group (Lieberman, 1967). Note that the primary difference between the fundamental frequency contour of this utterance and Figure 1 is that the fundamental frequency rises at the end of the breath-group. In some instances, a +breath-group ends with a level fundamental-frequency contour. The significant point is that it does not end with a falling fundamental-frequency contour. Similar acoustic and psychoacoustic data is again available (Chiba, 1935; Jassem, 1959; Fromkin and Ohala, 1968; Mattingly, 1966, 1968; Matsui et al., 1968; Ohala, 1970; Vanderslice, 1970; Lieberman et al., 1970). Armstrong and Ward (1926) and Jones (1932) also postulate similar acoustic correlates for Tune II as does Pike (1945) for the "tentative pause [\]." Trager and Smith (1951) postulate similar acoustic correlates for the "terminal junctures [\] and [\\]." They differentiate between the juncture [\] which ends with a level pitch contour and [\\] which ends with a rising fundamental-frequency contour. Note that these intonation contours are also true suprasegmentals (whether the transcription is in terms of Tune II, terminal juncture [\\], etc.). The fundamental frequency at the end of the contour is defined with respect to its behavior earlier in the contour. The listener must keep track of the entire contour in order to "decode" the final fundamental-frequency contour. Note that this automatically makes intonation contours into speech segmenting devices that have fairly long spans.

In contrast to the "long span" suprasegmental intonation contours other prosodic features have shorter spans. The feature which we shall call prominence (Lieberman, 1967, 1970; Lieberman et al., 1970) generally spans only a single syllable. Its acoustic correlates involve local increases in the fundamental frequency of phonation, the amplitude of the speech signal, and the duration of the segment (Fry, 1958; Jassem, 1959; Lieberman, 1960, 1967, 1970; Wang, 1962; Morton and Jassem, 1965; Fonagy, 1966; Ladefoged, 1969; Lehiste, 1961; Rigault, 1962; Hadding-Koch, 1961; Bolinger, 1958). The phonetic quality of the prominent syllable also may change (Lehiste and Peterson, 1959; Fry, 1965). The formant frequencies of the +prominent syllable show less coarticulation with adjacent segmental phonetic elements (Lindblom, 1963).
Acoustic and physiologic data for a marked, +breath-group (after Lieberman, 1967)
The acoustic correlates of prominence all seem to reflect increased activity of the muscles that are involved in speech production (Fonagy, 1966; Harris et al., 1968). All or some of the acoustic correlates can be used to mark a prominent syllable (Fry, 1958; Jassem, 1959; Lieberman, 1960). Note that the acoustic correlates that manifest prominence (excepting changes in formant frequencies) are again "prosodic" effects that interact with the long-span intonation to effect the total prosodic structure of an utterance.

Another phenomenon that appears to be a short-span prosodic feature is what we shall term accent. Whereas prominence appears to involve the synergetic activity of many muscles which act in concert to produce increases in the relative fundamental frequency, amplitude, and duration of a segment, an additional feature of accent exists. Accent appears to involve only contrasts in the fundamental-frequency contour. A syllable marked with the feature accent thus may have the acoustic correlate of a sudden decrease in fundamental frequency (Bolinger, 1958; Morton and Jassem, 1965; Katwijk and Govaert, 1967; Ohman, 1968; Barron, 1968; Vanderslice, 1970). We have deliberately avoided using the term "stress" which appears in many studies of the prosodic features since stress appears to be an abstract linguistic construct (Chomsky and Halle, 1966). In certain instances linguistically stressed syllables are not manifested phonetically by either prominence or accent (Lieberman, 1965; Barron, 1968). The feature of accent is also not always necessarily used to manifest underlying linguistic stress. The segmental "tone" systems that occur in so many languages may involve the feature (or features) of accent at the phonetic level (Ohman, 1968; Wang, 1967). It is important to continually bear in mind the dichotomy between the phonetic and semantic levels of language. Phonetic features in themselves have no inherent "meaning." The status of the term "stress" in generative linguistic studies, which unfortunately conflicts with its use in many phonetic studies, motivates our terminology.

Studies of the Larynx and Subglottal Vocal Mechanism

We have briefly discussed the acoustic correlates of some of the prosodic features that appear to have a reasonable basis in quantitative experimental evidence. There are other possibilities which need to be explored and we shall return to this topic. It is, however, necessary to first review the status of recent research on the anatomical mechanisms that

---

6There is no general agreement on terminology. Vanderslice (1970), for example, for similar reasons also avoids using the term "stress" at the phonetic level. He uses the term accent in about the same way as we do but he uses the term "emphasis" for the feature prominence, "cadence" for the acoustic correlates of the unmarked, -breath-group, and the term "endglide" for the acoustic correlates of the marked, +breath-group. Vanderslice's choice of different terminology appears to be, in part, motivated by theoretical differences concerning the articulatory maneuvers that underlie the acoustic correlates of these prosodic features as well as the status of the motor theory of perception. We will come to these aspects of modern theory later in our discussion. While theoretical differences exist, there appears to be, however, general agreement concerning the relevant acoustic phenomena.
appear to generate the acoustic correlates of the prosodic features. In order to understand the range of possible prosodic features we need to understand the constraints of the speech-production mechanism. Recent studies have fortunately made new information available that should be of benefit for further research on the prosodic features.

The acoustic theory of speech production shows that the acoustic speech signal can be regarded as the product of a source and a filter function (Chiba and Kajiyama, 1958; Fant, 1960). For voiced sounds, the source is the quasiperiodic series of "puffs" of air that exit from the larynx as the vocal cords rapidly move together and apart. The filter function is determined by the area function of the supralaryngeal vocal tract. As a first approximation, the glottal source and the supralaryngeal vocal tract do not interact. There are some interactions which we will discuss, but we can differentiate between the controlled variations of the supralaryngeal vocal tract's filter function and the controlled changes of the glottal source.

The fundamental frequency of phonation, which is the primary physical correlate of perceived pitch (Flanagan, 1965), is determined by the rate at which the vocal cords adduct and abduct. The energy content and amplitude of the glottal source, i.e., the glottal volume velocity waveform, which excites the supralaryngeal vocal tract is, in turn, determined by the rate at which the vocal cords move. When the vocal cords move faster and more abruptly, the glottal waveform more closely approximates a "pulse" (Timcke et al., 1958). The energy in the higher portions of the glottal spectrum is enhanced under these conditions (Flanagan, 1965). Since the fundamental frequency of a typical male speaker is about 120 Hz whereas the first and second formant frequencies of a vowel like /a/ are 700 and 1100 Hz (Fant, 1960), increasing the "high-frequency" energy content of the glottal source will increase the amplitude of the speech signal (Fant, 1960). Changes in fundamental frequency and amplitude are thus largely determined by the larynx for voiced sounds. It therefore is essential to know how the larynx functions in order to construct a viable phonetic theory.

The Myoelastic-Aerodynamic Theory of Phonation

In the early nineteenth century Johannes Müller (1848) first developed the concepts that have resulted in the myoelastic-aerodynamic theory of phonation (Van den Berg, 1958). This theory, which accounts for the known behavior of the larynx, states that the vocal cords (or vocal folds) rapidly move together and apart during phonation as a result of aerodynamic and aeroelastic forces. The vocal cords, in other words, passively move inwards and outwards as forces developed by the airstream rapidly alternate. The laryngeal muscles can adjust the initial position of the vocal cords, which determines whether phonation will or will not take place. The laryngeal muscles can also adjust the tension and mass of the vocal cords (Hirano & Ohala, 1969), which influences the manner in which the vocal cords move. The motive force for phonation is, however, provided by the airstream out from the lungs. The air pressure of the air in the lungs, which during phonation is nearly identical to the subglottal air pressure (Mead and Agostini, 1964), determines, in part, the rate at which the vocal cords move. Subglottal air pressure thus is an important factor in determining the fundamental frequency of phonation. The subglottal air pressure is itself a function of both the impedance of the glottis (i.e., the resistance which the larynx offers to the airflow)
and the force generated by the subglottal respiratory system. Since the glottal impedance is relatively high during phonation (Van den Berg, 1957) the activity of the subglottal respiratory system enters as a factor in the articulatory maneuvers that underlie the acoustic correlates of the prosodic features. The relative importance of the laryngeal muscles and the subglottal respiratory system to the control of fundamental frequency is a crucial factor in determining the articulatory implementation of the prosodic features. The investigation of this problem has shown that the larynx is a rather complex device. We will attempt to review some of the pertinent studies without becoming too involved in the physiology and anatomy of the larynx and the subglottal respiratory system.7

Air Pressure and Fundamental Frequency

The question which we shall review is beguilingly simple. An observer notes a change in the fundamental frequency of phonation during the production of a sustained note while a speaker is singing, or within the fundamental-frequency contour associated with a short utterance. The observer wants to know whether the change in f0 (fundamental frequency) follows from a change in the subglottal air pressure developed by the subglottal respiratory system or whether the f0 change follows from a change in the tension of the muscles of the larynx. The answer to this question appears to be that the larynx has many "modes" of phonation and that the effects of changes in the activity of the subglottal or laryngeal muscles on f0 are different in different modes of phonation.

The first quantitative experiments on the dynamics of the larynx involved the excitation of excised larynges in which the "lateral" tension on the vocal cords (Van den Berg, 1960) was changed by simulating the activity of the cricothyroid muscle (Müller, 1848). Müller in these experiments was able to simulate the activity of this muscle by an arrangement of pulleys, strings, and weights. He was able to change the force that this muscle would exert on the vocal cords while he simulated the flow of air out from the lungs by blowing air through a tube beneath the excised larynx. Müller found that three conditions had to be satisfied in order for phonation to take place. The vocal cords had to be moved inward from the open position that they assume for respiration and a minimum laryngeal tension and a minimum subglottal air pressure were necessary. Different combination of laryngeal tension and subglottal air pressure resulted in different fundamental frequencies of phonation. Müller could not precisely measure either the fundamental frequency of phonation (he subjectively matched pitch) or the subglottal air pressure, but modern refinements of this experiment have replicated his results.

7The reader can refer to Ladefoged et al. (1958), Mead and Agostini (1964), and Bouhuys et al. (1966) for detailed discussions of the mechanics of the subglottal system as well as techniques for the measurement of subglottal air pressure (Lieberman, 1968b). The myoelectric-aerodynamic theory of phonation was challenged by Husson (1950), who proposed that the muscles of the larynx provided the motive force for phonation. Husson claimed that the laryngeal muscles actively contracted in order to produce each fundamental period. The role played by subglottal pressure in determining the fundamental frequency of phonation was therefore minimal in Husson's erroneous theory (Van den Berg, 1958).
Van den Berg (1957, 1960) in a series of experiments with excised human larynges has found that there are two distinct "registers" in which phonation occurs. The laryngeal muscles, by adjusting the position and shape of the vocal cords, determine the register. In the chest register the vocal cords have a thick cross section, the part of the vocal cords that is set into motion is long, the tension applied by the cricothyroid muscle is relatively low, some "medial compression" applied by muscles like the interarytenoids and lateral cricoarytenoids is present, and the vocal cords collide in an "inelastic manner" as they move together in each fundamental period (Negus, 1949; Tincke et al., 1958; Van den Berg, 1957, 1960, 1968; Lieberman, 1967; Hirano and Ohala, 1969). In the "falsetto" register the vocal cords have a thin cross section, high lateral forces, a small vibrating mass, and elastic collisions, or no collisions, as the vocal cords move together in each fundamental period.

The vocal cords tend to move more abruptly in the chest register since their thick cross section (Hollien and Curtin, 1960) allows a nonlinear force, the Bernouilli force, to be realized as the air moving out from the lungs is forced through the glottal constriction (Van den Berg, 1957; Flanagan, 1965; Lieberman, 1967). The glottal excitation therefore has more high-frequency energy for phonation in the chest register. The sensitivity of the larynx to changes in subglottal air pressure also varies for the two registers. Van den Berg (1960) in experiments with excised human larynges found that the sensitivity of the larynx to changes in subglottal air pressure ranged from 2.5 Hz to 20 Hz/cm H$_2$O. In these experiments the tension and configuration of the vocal cords of an excised larynx were held constant while the subglottal air pressure was changed. The tensions and positions of the vocal cords were then set to another set of parameters and the subglottal air pressure was again varied. In this manner a number of plots relating fundamental frequency and subglottal air pressure were obtained. The only criticism that can be leveled at experiments of this sort is whether the air pressures, tension, and positions that the experimenter imposes on the vocal cords of the excised larynges are typical of those that occur in vivo. Experiments in which speakers are asked to sing notes while the subglottal air pressure is artificially and abruptly changed indicate that Van den Berg's results are probably realistic.

Lieberman, Knudson, and Mead (1969), for example, performed an experiment in which a single speaker attempted to sing sustained notes while sinusoidal modulations in air pressure were imposed on his subglottal air pressure. The singer sang at both "loud" and "soft" levels at different pitch levels. The rate of change of fundamental frequency was found to vary between 2 and 20 Hz/cm H$_2$O when the fundamental-frequency variations of the sustained notes were correlated with the sinusoidal air-pressure modulations. The speaker in this experiment did not attempt to match any reference tones as he sang.

A number of experiments have been conducted where a speaker sings a note while his chest is abruptly pushed inward by a light push. This causes his subglottal air pressure to abruptly rise. Low rates of change of $f_0$ with respect to air pressure (from 2.0 to 5.0 Hz/cm H$_2$O) have been reported by Ladefoged (1962), Ohman (1968), and Fromkin and Ohala (1968) using this technique. In a recent experiment (Ohala and Ladefoged, 1970) somewhat
higher rates of change of \( f_0 \) with respect to subglottal air pressure variation (to 10 Hz/cm H\(_2\)O) have been reported using this technique. Other studies have correlated fundamental-frequency changes with subglottal air-pressure variations during the production of short sentences. Lieberman (1967) obtained a value of about 17 Hz/cm H\(_2\)O by this technique. The activity of the laryngeal muscles was, however, not monitored in this study and some of the variations in \( f_0 \) that were ascribed to subglottal air-pressure variations may have been due to the activity of laryngeal muscles tensing. An analysis of the data of Fromkin and Ohala (1968), in which the activity of several laryngeal muscles was monitored, however, shows that the rate of change of \( f_0 \) with respect to subglottal air-pressure variation is 12.5 Hz/cm H\(_2\)O (Lieberman et al., 1970). It is interesting to note that this same speaker apparently showed less sensitivity to variations in subglottal air pressure (about 5 Hz/cm H\(_2\)O) when his activity was monitored while he sang.

Some of the differences between the rates of change of \( f_0 \) with respect to subglottal air pressure that we have discussed may be perhaps ascribed to experimental artifacts that arise in the measurement of subglottal air pressure. The physiology of the respiratory system can make the measurement of subglottal air pressure from measurements derived from esophageal balloons (Ladefoged, 1969) rather involved (Bouhuys et al., 1966; Lieberman, 1968). Nonetheless it is clear that the sensitivity of the larynx to changes in subglottal air pressure is variable. The theoretical laryngeal model developed by Flanagan and Landgraf (1968) and Flanagan (1968) predicts that the variation of \( f_0 \) with subglottal air pressure will be different for different "modes" and different "registers" of phonation. This model indicates that the larynx is least sensitive to variations in air pressure for phonation in the chest register that involves inelastic collision of the vocal cords. The larynx is most sensitive to air pressure for falsetto register phonation with elastic collision. The total range of variation that the model predicts is 2 to 20 Hz/cm H\(_2\)O.

The experimental data on excised larynges as well as data derived from both sustained "singing" and speech in humans thus supports this laryngeal model. We can tentatively conclude that the larynx can either be sensitive or insensitive to variations in subglottal air pressure. The laryngeal configurations used by trained singers probably result in minimum sensitivity since this would simplify the pitch-control problem. The larynx assumes a different posture in singing (Sundberg, 1969). In singing, controlled variations in \( f_0 \) probably are the consequence of laryngeal maneuvers. In speech production, speakers apparently may use laryngeal configurations that result in a relatively high sensitivity of \( f_0 \) to air-pressure variations (Lieberman, 1967; Lieberman et al., 1970; Kumar and Ojamaa, 1970).

Although it would be far "simpler" if all changes in \( f_0 \) during speech were exclusively due to laryngeal maneuvers (Vanderslice, 1967, 1970; Fromkin and Ohala, 1968; Ohman, 1968; Ohala, 1970), this does not appear to be so. The data reported by Fromkin and Ohala (1968), which forms the substantive base of the claim that laryngeal maneuvers exclusively generate the controlled \( f_0 \) changes of intonation, indeed indicates that both subglottal pressure changes and laryngeal maneuvers must be taken into account (Lieberman et al., 1970).
"Uncontrolled" f0 Variations

The theoretical model of the larynx that we have noted (Flanagan and Landgraf, 1968; Flanagan, 1968) also explains other aspects of f0 variation during speech. Peterson and Barney (1952) in their study of vowel formant frequencies note that different vowels appear to have slightly higher or lower average fundamental frequencies. Similar results have since been noted by House and Fairbanks (1953), Lehiste and Peterson (1959), and Swigart and Takefuta (1968). Vowels that have low first-formant frequencies, e.g., /u/ and /i/, have higher fundamental frequencies. Vowels that have high first-formant frequencies, e.g., /a/, have lower fundamental frequencies. The Flanagan and Landgraf (1968) model of the larynx indicates that these effects, which involve offsets of about 10 Hz for an average f0 of 120 Hz, are due to aerodynamic coupling between the supralaryngeal vocal tract and the larynx. The presumed independence of the glottal source and the supralaryngeal vocal tract is, as we noted, only a first approximation.

The model developed by Flanagan (1968) also indicates that the sensitivity of the larynx to changes in subglottal air pressure will be affected by the supralaryngeal vocal-tract configuration. The highest rates of change of f0 with respect to subglottal air pressure occur for vowels with low first-formant frequencies. The model predicts that the vowel /a/ which has a high first-formant frequency will result in the lowest variations in f0 with changes in subglottal air pressure, all other parameters being equal. These variations may, in part, account for some of the different values for the sensitivity of f0 to subglottal air-pressure variation that have been obtained for sung vowels, where /a/ is usually produced, and speech.

We have been oversimplifying our discussion of the relationship between subglottal air pressure and fundamental frequency. The transglottal air pressure is actually the factor that we should keep in mind when we discuss these variations (Flanagan, 1965; Lieberman, 1967). We have implicitly assumed that the supraglottal air pressure stays constant while the subglottal air pressure varies. This is, of course, true in studies where a speaker sings a single, sustained vowel. In the production of connected speech this is, however, not the case, as the oral air pressure abruptly builds up during the production of stops like /b/ where the lips close, as well as, to a lesser degree, for other consonantal sounds. These variations in transglottal air pressure will cause variations in f0 that are concomitant with segmental phonetic elements (Ohman, 1968; Lieberman et al., 1970).

Still other variations in f0 will arise from coarticulation phenomena (Lindblom, 1963). The larynx is continually reset from its closed phonation position to more open positions for the production of unvoiced segmental phonemes. These transitions are comparatively slow. It takes almost 100 msec to move the vocal cords from an unvoiced to a voiced configuration (Lieberman, 1967; Lisker et al., 1969; Sawashima et al., ms.). Variations in fundamental frequency are thus to be expected as the vocal cords either finish a closing maneuver or begin to anticipate an opening maneuver.
These perturbations of the $f_0$ contour can be correlated with voiced and unvoiced stops (House and Fairbanks, 1953; Ojamaa et al., 1970). They may be secondary cues for the perception of voiced and unvoiced stops (Haggard, 1969).

The larynx is obviously not an isolated appendage of the human body. It is quite possible that gross skeletal maneuvers can affect the configuration and the tension of the vocal cords through the complex system of ligament, cartilage, and muscle that connects the larynx with the skeletal frame (Sonninen, 1956). Some of the data that traditionally has been cited as supporting evidence for the mechanical interactions affecting $f_0$ should perhaps be reappraised in the light of the aerodynamic interactions that can occur between the larynx and the supralaryngeal vocal tract. The higher $f_0$'s associated with /u/ and /i/ are sometimes interpreted as evidence for mechanical interaction between the muscles of the tongue and the larynx (Ohala and Ladefoged, 1970). The higher $f_0$ of these vowels, however, appears to be an aerodynamic effect. Radiographic data (Perkell, 1969) of the vowels makes it difficult to see why both of these vowels should have higher $f_0$'s from mechanical interactions while the vowel /a/ has a lower $f_0$.

Ohala and Hirano (1967) and Ohala and Hirose (1969) have obtained electromyographic data that shows that the sternohyoid muscle is active when a speaker sings at either high or extremely low fundamental frequencies. Whether similar mechanisms play a role during connected speech is still a question. Ohman (1968) has suggested that such maneuvers may underlie a proposed phonetic feature of -accent which results in an $f_0$ fall. Ohala (1970) presents data that shows increased sternohyoid activity accompanying falling $f_0$ contours, but these $f_0$ changes could also be the consequence of either falling subglottal air pressure or of the opening of the larynx in anticipation of an unvoiced stop.

Some of the distinctions that have been occasionally made between subglottal air pressure and laryngeal maneuvers are meaningless. Subglottal air pressure is the consequence of both the impedance offered by the larynx when the vocal cords are in their adducted or nearly adducted phonation position (Lieberman, 1967) and the activity of the subglottal respiratory system which forces air out from the lungs. Before an unvoiced segment, subglottal air pressure will fall. This will affect the $f_0$ contour (House and Fairbanks, 1953; Ojamaa et al., 1970). The activity of the larynx that causes this effect is, however, an articulatory manifestation of the segmental phonetic feature -voiced (Chomsky and Halle, 1968) rather than an articulatory manifestation of a prosodic feature. In other instances (Ladefoged et al., 1958; Ladefoged, 1962, 1968, 1969; Lieberman et al., 1970), changes in subglottal air pressure can be observed that clearly are the consequence of maneuvers of the subglottal respiratory system.

**Phonetic Theories**

We have presented a brief review of some of the factors that cause uncontrolled variations in $f_0$. These phenomena perhaps explain why many phoneticians have been reluctant to work with "objective," electronically derived, fundamental-frequency contours. These contours frequently contain many errors since it is extremely difficult to derive fundamental frequency by means
of electronic devices (Flanagan, 1965). However, even when the electronic
devices work, they show minute variations in fundamental frequency that are
not acoustic correlates of prosodic features. The human observer will not
pay any attention to these variations in the framework of the prosodic fea-
ture system. He may perhaps use some of these variations as secondary cues
for the perception of the segmental phonetic elements that generate them
(Lehiste and Peterson, 1959; Haggard, 1969). Some of the variations in $f_0$
that occur in speech may simply be the result of chance variations in laryn-
geal muscle tension or the activity of the subglottal respiratory system.
Speakers do not appear to take great care in producing exactly the same $f_0$
contour for the same utterance (Lieberman, 1967). When the $f_0$
contours of the "same" sentence are compared for several speakers, startling
differences can be seen (Lieberman, 1965, 1967; Rabiner et al., 1969). The utterances,
of course, may not be the "same" at the emotional, "level 4" aspect of com-
munication. Contours that have rather different "fine structures" with
respect to their $f_0$ variations do not appear to have any linguistic import.
Listeners are unable to differentiate the contours at any linguistic level
though they may ascribe different emotional contexts to the contours
(Lieberman and Michaels, 1962).

Phonetic theories like that of Isacenko and Schadlich (1963) are un-
convincing since they rely on small 5 to 10 Hz variations over long intona-
tion contours in synthesized signals. Linguistic studies like Bierwisch
(1966) that attempt to derive "grammatical rules" that will specify intona-
tion contours in this detail are thus overspecified. Preliminary perceptu-
tal experiments with synthesized speech appear to show that only a few
gross factors are important when the $f_0$ contour that accompanies an ut-
terance is specified; the direction of the terminal contour (whether it
falls or rises) and the point of the major prominence (if any occurs) of
the utterance must be specified. Listeners will accept almost all other
variations in the $f_0$ contour. They are, however, sensitive to the duration
of each segmental phonetic element with respect to its position in the ut-
terance (Mattingly, 1966, 1968; Matsui et al., 1968). Some phenomena that
have traditionally been associated with $f_0$ in the detailed transcriptions
like those of Kingdon (1958), Schubiger (1958), and Halliday (1967) may
perhaps have durational correlates.

The perception of intonation by linguists often appears to be "con-
taminated" by their analysis of other levels of language. In an experiment
on the perception of intonation (Lieberman, 1965) linguists trained in the
Trager-Smith (1951) notation were asked to transcribe a set of eight ut-
terances. Electronic processing equipment was available that abstracted
the acoustic correlates of the prosodic features that the linguists were
ostensibly transcribing from the words of the message. When the linguists
were presented with these isolated prosodic contours, which modulated a
fixed vowel, they were unable to transcribe the pitch levels and junctures
that they noted on hearing the complete utterance. The linguists' prosodic
transcriptions were, moreover, more accurate when they listened to the iso-
lated acoustic features of fundamental frequency and acoustic amplitude as
functions of time. The pitch levels and junctures of the Trager-Smith nota-
tion apparently depended on the linguistic information conveyed by the words
of the message. The Trager-Smith notation appears to be a device whereby
semantic differences carried by different underlying phrase markers can be
recorded (Lieberman, 1967). The linguist using this system "hears" the
pitch levels that transcribe the "pitch morpheme" that he wishes to present. The notation of "pitch morphemes" follows from the "attitudinal meanings" that Pike (1945) states are conveyed by the prosodic features. Pike is essentially discussing some of the nonlinguistic, "level 4" functions of the prosodic features. Trager and Smith invent pitch morphemes that convey the semantic information that they are unable to account for with the immediate constituent grammar that is the basis of their linguistic theory (Postal, 1964).

The Motor Theory of Speech Perception

One of the most influential developments of recent years is the modern version of the motor theory of speech perception (Liberman et al., 1967). The motor theory essentially states that the constraints imposed by the human speech production are structured into a "speech perception mode." The motor theory thus also states that speech is perceived in a different manner than other acoustic signals. There is a large body of experimental evidence that supports this theory though many points still are in dispute. It is important to note that the motor theory does not state that listeners must "learn" to decode speech through an explicit knowledge of the process of speech production. It thus is irrelevant that listeners who are unable to speak can perceive speech (Lenneberg, 1967). There are undoubtedly special neural detectors in man that are structured in terms of the sounds that the human speech apparatus makes. Similar neural devices have been found in frog (Capranica, 1965) and in cat (Whitfield, 1969). The general motor theory is reviewed in this volume by M. Studdert-Kennedy. It is of interest here since it accounts for some otherwise perplexing phenomena in the perception of the prosodic features.

Several studies have noted that the stressed syllables of English bisyllabic words like rebel and rebel may receive the phonetic feature that we have called +prominence when they occur in isolation. One of the principal articulatory maneuvers that underlies the acoustic correlates of +prominence is a momentary increase in subglottal air pressure. The listener's responses to the stressed syllables indicate that he appears to be responding to the magnitude of the subglottal air-pressure peak (Ladefoged, 1962, 1968, 1969; Ladefoged and McKinney, 1963; Lehiste and Peterson, 1959; Fonagy, 1966; Lieberman, 1967). Jones (1932) is perhaps the modern source.

When lists of isolated words that have contrasting stress patterns are read in phonetic experiments, the readers will use the feature +prominence to manifest primary stress. The stress levels generated by the rules of the phonologic component of the grammar (Chomsky and Halle, 1968) do not, however, appear to be manifested by the feature prominence in normal discourse. It is necessary to construct an utterance in which semantic information is carried by a "morpheme" of "emphasis" or "focus" in the underlying phrase marker (Chomsky, 1968; Postal, 1968). These morphemes may be ultimately realized phonetically in different ways. The ultimate result of certain optional transformations and phonologic rules may, for example, yield the sentence, Tom did run home. Other transformations and phonologic rules could yield, Tom RAN home. (where the capitals indicate the presence of +prominence at the phonetic level). In other words, the occurrence of the feature +prominence usually seems to be attributable to the presence of some morpheme in the underlying phonologic component of the grammar.
for this theory for the perception of +prominence. (Jones uses the term stress where we use prominence.) The motor theory of speech perception is most illuminating when we consider the interactions that can occur between +prominent syllables and the sentence intonation's \( f_0 \) contour. In Figure 3, two fundamental frequency contours are schematized. Both contours were used with a standard carrier phrase in a controlled psychoacoustic experiment (Hadding-Koch and Studdert-Kennedy, 1964). The listeners in this experiment said that both of these contours were questions that had the same terminal rise. Note that the two contours in Figure 3 have different terminal \( f_0 \) rises. The listeners were not merely responding to the physically present fundamental-frequency signal. They instead appeared to be evaluating the terminal fundamental-frequency contours in terms of the underlying articulatory maneuvers that would be present in natural speech.

The data presented by Lieberman (1967), Fromkin and Ohala (1968), Ohala (1970), and Lieberman et al. (1970) all show that yes-no questions in English are produced with +breath-groups where the tension of certain laryngeal muscles increases at the end of the breath-group to effect a "not falling" \( f_0 \) contour. The +breath-group, which is used for short unemphasized declarative sentences in English, ends with a falling terminal \( f_0 \) contour that is a consequence of the falling subglottal air pressure that occurs at the breath-group's end (Lieberman, 1967; Fromkin and Ohala, 1968; Ohala, 1970; Lieberman et al., 1970). The data in Figures 1 and 2 show examples of a -breath-group and a +breath-group. Note that the +breath-group also ends with a falling subglottal pressure function. The increased tension of the laryngeal muscles counters the falling subglottal air pressure to produce a rising terminal \( f_0 \) contour in Figure 2.

We have noted that the articulatory maneuvers that underlie the prosodic feature +prominence include a momentary increase in subglottal air pressure. The momentary increase in subglottal air pressure will produce an increase in the fundamental frequency since the fundamental frequency of phonation is a function of both laryngeal tension and transglottal air pressure.

The listeners in the Hadding-Koch and Studdert-Kennedy experiment appear to be evaluating the intonation contours in Figure 3 in terms of their "knowledge" of the articulatory bases of these features. They "know" that rising terminal \( f_0 \) contours must be the consequence of increased laryngeal tension. They "know" that the nonterminal \( f_0 \) peak is a consequence of a peak in subglottal air pressure. They also appear to "know" another fact about speech production that we have not yet discussed. Nonterminal peaks in the subglottal air pressure function will result in a lower subglottal air pressure after they occur. This effect can be seen in the data of Lieberman (1967) as well as in the independent data of Fromkin and Ohala (1968) which is discussed in Lieberman et al. (1970). This "air-pressure perturbation" effect appears to be a consequence of the physiology of the lungs. The elastic recoil of the lungs and the respiratory muscles which regulate subglottal pressure (Mead and Agostini, 1964) appear to be "programmed" in terms of the overall breath-group, whereas independent instructions result in the pressure peak associated with +prominence. The +prominence pressure peak lowers the volume of the lungs which, in turn, lowers the air pressure generated by the elastic recoil (Bouhuys et al., 1966; Lieberman, 1967:54,71,98-100; Lieberman et al., 1970; Kumar and Ojamaa, 1970).
Synthesized intonation contours that listeners perceived to have the "same" terminal pitch contours (from data of Hadding-Koch and Studdert-Kennedy, 1964)

Fig. 3
The "articulatory decoding" that the listeners use in perceiving the terminal rises in Figure 3 thus goes as follows: the lower contour's prominence \( f_0 \) peak is 370 Hz, whereas the upper contour's is 310 Hz; since a higher \( f_0 \) implies the presence of a greater subglottal air pressure peak, the listener knows that the lower contour employed a greater nonterminal subglottal air pressure peak; the listener now "knows" that the terminal subglottal air pressure of the lower contour is lower than the upper contour's. This follows from the air pressure perturbation effect. An equivalent degree of laryngeal tension would thus result in a lower terminal fundamental frequency. The listener thus evaluates the terminal fundamental-frequency contours in terms of the laryngeal tension that would be present in natural speech.

It is important to remember that this "motor theory perception" of intonation does not imply that the listeners actually "know" at a conscious level any of the relationships that we have discussed. People "know" many complex relationships at some neural level without any conscious knowledge of the fact. The act of respiration itself involves many "reflexes" and "regulatory mechanisms" that are not part of the conscious knowledge of the casual "breather." We all must breath but few of us are aware of the maneuvers that we must employ to generate a relatively steady subglottal air pressure over the nonterminal portions of an utterance (c.f., Figs. 1 and 2). We nonetheless "know" how to regulate the variable pressure that would be generated by the elastic recoil of our lungs (Mead and Agostini, 1964; Ladefoged et al., 1958). The "knowledge" of these aspects of speech production probably are the result of a long evolutionary process in man. Modern man's hominid ancestors gradually acquired the anatomical mechanisms that are involved in human speech (Lieberman and Crelin, 1971). The neural decoding mechanism or neural detectors that are involved in the decoding of speech are thus part of man's evolutionary endowment, and we should not really be surprised to find relatively complex pattern detectors. Similar detectors have been found in frogs and cats (Capranica, 1965; Whitfield, 1969). Birds probably employ similar ones (Greenewalt, 1968), and we are presumably not less well endowed.

The point of view that we have explicated, i.e., the structure that is imposed on the production and the perception of speech by the anatomical and physiologic constraints of the human speaker, also is consistent with the intonational forms that human languages tend to use most. If we think of the simplest way in which a speaker can regulate subglottal air pressure to speak, we arrive at the intonation pattern that is most common in the languages of the world (Chiba, 1935; Hadding-Koch, 1961; Revtova, 1965; Lieberman, 1967). This pattern, which involves a terminal falling \( f_0 \) contour, follows from the simplest, or "archtypal," pattern of muscular control that can be used to generate a relatively steady subglottal air-pressure contour over the course of an utterance. The simplest pattern obviously involves a state of minimal laryngeal control over the course of the utterance. At the end of the utterance, the fundamental frequency of phonation will rapidly fall since the subglottal air pressure must change from a positive to a negative air pressure. A negative air pressure is the natural consequence of the requirements of respiration. The speaker has to get air into his lungs in order to breathe. The vegetative process of respiration thus determines the form of what we have called the unmarked, or normal, breath-group.
(Fig. 1). All people and all languages obviously do not use the simplest "un-
marked" state (Chomsky and Halle, 1968). The +breath-group, which is "marked,"
obviously involves a more complex pattern of muscular activity where the
laryngeal muscles must be tensioned at the end of the breath-group in order
to counter the falling f0 contour that would result from the "vegetative"
fall in subglottal pressure.

Different languages and different speakers also probably do not make
use of the "archtypal," unmarked breath-group even when they retain the
falling terminal f0 contour for most utterances. Idiosyncratic and language-
specific patterns of laryngeal activity are often used to produce f0 varia-
tions throughout the nonterminal portions of the breath-group (Lieberman,
1967; Fromkin and Ohala, 1968; Lieberman et al., 1970; Kumar and Ojamaa,
1970). The situation is no different from the implementation of other phone-
tic features. Idiosyncratic and language-specific modifications also occur
(Lieberman, 1970).

We have noted that a number of studies (Bolinger, 1958; Hadding-Koch,
1961; Fromkin and Ohala, 1968; Morton and Jassem, 1965; Barron, 1968;
Vanderslice, 1967, 1970; Ohala, 1970) indicate that linguistic stress can
be manifested by sudden falls in f0. These sudden falls in f0 could be im-
plemented either by tensing laryngeal muscles that would cause fundamental
frequency to fall or by relaxing laryngeal muscles that raise f0 when they
tense. There is no a priori reason to assume that all phonetic features
must be implemented by tensing a particular muscle. Phonetic features do
not stand in a one-to-one mapping with particular muscles. Even when there
is a close relation between a particular phonetic feature and a muscle,
e.g., nasality and the levator palatini muscle which controls the velo-
pharyngeal port, the implementation of the feature may involve relaxing the
muscle in question (Lieberman, 1970). It is therefore possible that abrupt
falls in f0 could be implemented by relaxing muscles that in their tensed
state maintain a higher f0. Either the cricothyroid muscle which applies
lateral tension to the vocal cords or the muscles that adduct the vocal
cords and apply medial compression (Van den Berg, 1960) could do this.

Several studies (Vanderslice, 1967; Fromkin and Ohala, 1968; Ohala,
1970) have stated that these abrupt falls in f0 are the consequence of
tensing the sternohyoid muscle. This muscle can alter the vertical posi-
tion of the larynx and a lower laryngeal position is said to produce a
lower fundamental frequency (Vanderslice, 1967). The data in these studies
show that the nonterminal falling f0 contours are sometimes the consequence
of some laryngeal maneuver. (Falling f0 is seen where the subglottal pres-
sure is steady.) The activity of the sternohyoid muscle, however, does
not appear to be related to these f0 falls. The sternohyoid muscle has
been shown to be active in the maneuvers that singers use to sing at both
low and high fundamental frequencies (Hirano and Ohala, 1969; Ohala and
Hirose, 1969). The data of Ohala and Hirose (1969) however show that "any
gesture of speech or nonspeech that would most likely require a lowering or
fixation of the hyoid bone tended to show an increase in the activity of
this muscle." The position of the larynx and hyoid bone do change during
connected speech. This motion, however, appears to be an articulatory
maneuver that is directed at lowering the formant frequencies of segmental
phonetic elements (Perkell, 1969) as well as the sustenion of phonation in
+voiced stops like /b/. The total volume of the supralaryngeal vocal tract
expands during the closure interval of these stops in order to sustain voicing (Perkell, 1969). The larynx also may be lowered at the end of phonation as part of the general adjustment of the larynx for inspiration. When the vertical position of the larynx is correlated with the average fundamental frequencies of the vowels (Peterson and Barney, 1952; Swigart and Takefuta, 1968; Perkell, 1969; Sundberg, 1969) it is apparent that f₀ is not correlated with larynx height. The highest and lowest larynx heights, /I/ and /u/, have the highest average fundamental frequencies. The average fundamental frequencies of these vowels are apparently due to aerodynamic coupling with the larynx (Flanagan and Landgraf, 1968) rather than larynx height adjustment.

The emphasis on showing that tensing a laryngeal muscle will lower f₀ in the Vanderslice (1967), Fromkin and Ohala (1968), and Ohala (1970) studies follows from their belief that the subglottal air pressure contour plays virtually no part in determining the f₀ contour. The falling f₀ contour that occurs at the end of a -breath-group thus must have a laryngeal correlate. Since there is no evidence in the electromyographic data of Fromkin and Ohala (1968) that the laryngeal muscles that maintain f₀ relax, the presence of a laryngeal muscle that lowers f₀ as it is tensed is postulated. The close correlation that is manifested between the f₀ and subglottal air-pressure contours for the -breath-groups (the unemphasized declarative sentences) in the data of Fromkin and Ohala (1968) and Ohala (1970) is felt by these authors to be fortuitous. Ohala and Ladefoged (1970) recently have, however, found that air-pressure variations can change fundamental frequency at rates as high as 10 Hz/cm H₂O, whereas they earlier believed that the highest rate that occurred in speech was about 3 Hz/cm H₂O. The controversy regarding the sensitivity of the fundamental frequency to subglottal air-pressure variations, as we noted earlier, apparently reflects the fundamental character of the laryngeal source. The fundamental frequency of phonation can be insensitive to air-pressure variations as it is in the "modes" of phonation that appear to be used in singing. If the sensitivity of the larynx is measured in this mode of phonation, air pressure will have a negligible effect. The larynx, however, can also phonate in "modes" that make it sensitive to air-pressure variations. The data of Fromkin and Ohala (1968) and Ohala (1970), despite the claims made by these authors, show that the larynx is sensitive to air-pressure variations during the production of speech at a rate of about 12.5 Hz/cm H₂O.

Öhman (1968) in a study of the accent system of Swedish has accounted for a set of rather complex variations by means of two prosodic features which we shall call accent down and accent up. These features appear to be sufficient to generate the falling f₀ contours that manifest phonetic stress in English (Morton and Jassem, 1965; Barron, 1968) as well as some of the phenomena discussed by Vanderslice (1967, 1970), Fromkin and Ohala (1968), and Ohala (1970). These features would, however, not be necessary to account for the falling terminal f₀ contour of the archetypal normal -breath-group (-breath-group) which follows from the falling subglottal air-pressure contour. 9) Bolinger (1958) described a range of phenomena in

---

9In some instances, adult speakers can be observed who start to move their larynx towards its open respiratory configuration before the end of a -breath-group. The laryngeal muscles will, as they abduct the vocal cords,
English that also can be accounted for by means of these features. We are following Bolinger's precedent in our terminology. The implementation of \textit{accent up} would involve a laryngeal maneuver that raised $f_0$, whereas the implementation of \textit{accent down} would involve a laryngeal maneuver that lowered $f_0$. Note that in contrast to the feature \textit{prominence}, which can involve an increase in laryngeal tension that raises $f_0$ (Fromkin and Ohala, 1968; Harris et al., 1968; Lieberman et al., 1970; Lieberman, 1970; Ohala, 1970) the feature \textit{accent up} would only involve activity in the larynx. Its articulatory implementation would thus be more localized than the feature \textit{prominence} which appears to involve heightened muscular activity throughout the vocal tract (Fonagy, 1966; Harris et al., 1968).

The two features \textit{accent up} and \textit{accent down} may perhaps be the phonetic manifestations of the segmental "tones" that have been noted in many languages (Chang, 1958; Abramson, 1962; Wang, 1957). These tones clearly interact with the $f_0$ contour of the sentence, and they appear to be independent of \textit{prominence} (Chang, 1958). The intonational transcriptions developed by Jassem (1952), Lee (1956), Kingdon (1958), and Halliday (1967), which are based on their perception of "meaningful" prosodic events, may also reflect these two features of accent.

The studies of Bolinger (1958, 1961), Vanderslice and Pierson (1967), Nash (1967), and Crystal (1969), as well as those of Kingdon, Halliday, Schubiger, and Hadding-Koch, which we have cited, indicate that other prosodic features that are imposed on the \textit{breath-group} also must be investigated. Armstrong and Ward (1926) and Jones (1932) in their perceptually based studies, for example, note the presence of an element of \textit{emphasis} which may extend over an entire breath-group (a "tune" in their notation). It is not clear at this time whether this involves an additional feature or whether the scope of the feature \textit{prominence} can be extended over an entire breath-group. In like manner, Bolinger (1958, 1961) establishes a convincing case for features that result in gradual changes of $f_0$ over a comparatively long part of a breath-group. It is again not clear whether

lower $f_0$ by increasing the vibrating mass, lowering the tension of the vocal cords, and lowering the subglottal air pressure by reducing the glottal impedance (Lieberman, 1967). These premature opening maneuvers probably should be regarded as idiosyncratic variations on the archetypal form of the \textit{breath-group} (Lieberman, 1970) rather than as implementations of the feature \textit{accent down}. They are in the same class as the variations in voicing onset observed in the production of stops by Lisker and Abramson (1964) for isolated individual speakers of English. Other speakers, e.g., Speaker 1 in Lieberman (1967) and the speaker in Fromkin and Ohala (1968), do not open their larynges until the end of phonation in a \textit{breath-group}. If we classify idiosyncratic implementations of a feature as manifestations of other features, we would have to conclude that individual speakers had phonetic components that made use of different feature complexes.

Note that a premature opening of the larynx near the end of a \textit{breath-group} merely emphasizes the falling $f_0$ contour that usually is the consequence of the falling subglottal air pressure. The articulatory gesture of opening the larynx is not in opposition to the falling subglottal air pressure. It enhances the trend already established. Phonetic features must act distinctively.

201
these "ramps" should be regarded as the result of many small steps of +accent down or +accent up (depending on the direction of the $f_0$ contour) or whether new features should be introduced. The electromyographic techniques that are described by K. S. Harris in this volume will undoubtedly prove useful in resolving these and other questions. The introduction of physiologic and acoustic techniques has provided a reasonable advance over the theory presented by Stetson (1951). The introduction of new or refined techniques that allow the activity of muscles to be correlated with articulatory, physiologic, and acoustic data promises similar advances in the near future.

**Multileveled Versus Binary Features**

A problem that is perhaps more susceptible to controlled psychoacoustic experiments is whether prosodic features are multivalued or binary. This problem is, of course, not simply confined to the prosodic phonetic features. It has been raised many times with regard to the segmental phonetic features. The question can be partially resolved in terms of the mechanism available to the phonetic component of a grammatical theory. If phonetic features stand in a close relation to the acoustic signal (Jakobson et al., 1952) or to individual muscles or muscle groups (Ladefoged, 1967; Chomsky and Halle, 1968; Fromkin, 1968), multivalued features will have to be introduced.

If phonetic features are instead regarded as "state functions" at the articulatory level rather than as specific, invariant, muscular commands or articulatory maneuvers, a fairly powerful phonetic component must be introduced into the grammar to yield the actual articulatory maneuvers that underlie speech (Lieberman, 1970). A state function is, for example, the feature consonantal which does not involve a particular muscle or articulatory maneuver. The phonetic component of the grammar will involve "implementation rules" that can generate multivalued acoustic or articulatory phenomena from an input ensemble of binary phonetic features. It therefore is possible to have multivalued phenomena like the stress levels postulated by Trager and Smith (1951) even though binary phonetic features form the input to the phonetic component. The phenomenon that Bolinger and Gerstman (1957) and Lieberman (1967) termed "disjunctive" together with the features of prominence, accent down and accent up could, in theory, combine to map out a multivalued stress system. Disjunctive has the acoustic correlate of an unfilled pause. It therefore could combine with any or all of the other prosodic features that we have discussed to provide a distinct physical basis for multivalued stress levels.

Perceptual studies by Bolinger and Gerstman (1957), Lieberman (1965), and Barron (1968) suggest, however, that human listeners cannot differentiate more than two levels of stress in connected speech. Hadding-Koch (1961) attempts to correlate perceived stress levels in the Trager-Smith system with acoustic measurements of Swedish discourse, but the results are

---

10 The author is inclined to speculate that the case of emphasis extending over an entire breath-group can be treated as a special case of +prominence where the scope is the entire breath-group. C. J. Bailey (pers. comm.) has developed some convincing grammatical arguments for this approach. The gradual contours described by Bolinger, however, would appear to involve additional prosodic features.
not conclusive. In certain restricted contexts (Hart and van Katwijk, 1969) multivalued stress decisions can be made by trained listeners. These effects, as Hart and van Katwijk point out, may be experimental artifacts. The particular choice of features that is used to map out the stress levels does not appear to be a factor in these experiments, and a tentative conclusion is that only two levels of stress can be mapped out by the prosodic features. The process of vowel reduction which appears to be a consequence of the stress-assignment rules in English (Chomsky and Halle, 1968) can provide a physical basis for a third level of stress. It is important to remember that we are using the term "stress" here to signify the linguistically determined "level" that the rules of the phonologic component assign to a vowel. We are not using the term to signify a phonetic feature. A listener thus can perceive the stress levels of an utterance by means of internal computations that involve his knowledge of the phonologic stress assignment rules of his language and the constituent structure of the utterance. There need be no acoustic or phonetic events that specifically map out the stress levels (Lieberman, 1965; Barron, 1968).

In closing we must take note of one of our prefatory remarks. We have deliberately neglected many studies of intonation that are based on the unaided senses of a trained observer who transcribes what appear to be meaningful prosodic events. We have instead concentrated on recent studies that are based on quantitative acoustic, anatomical, and physiologic data as well as psychoacoustic data. This does not imply that we believe that auditorily based studies are useless. They bring to light many phenomena that would be overlooked in the small data ensembles that generally form the basis of more "quantitative" studies. We have also taken the liberty of making some arbitrary definitions concerning particular phonetic features that may not always agree with the definitions of these terms in other studies. Some degree of arbitrarity is, however, necessary in this regard in light of the differences in theory and method that differentiate various studies. The reader can, it is hoped, make the necessary name changes if he finds them desirable. The situation has hitherto been rather anarchic and we have been forced to bring some degree of order that may sometimes appear arbitrary by redefining several terms. The reader also will note that we have not ventured into the area that we termed "level 4" functions of prosody, e.g., Halliday (1967) and Crystal (1969). As linguistic theory develops, many of these functions will doubtlessly be amenable to linguistic analysis, but, as Crystal himself (1969a) notes, "before any attempt to integrate intonation with the rest of a description is likely to succeed, various preliminaries have to be gone through....One cannot assume that everyone means the same thing by such labels as 'confirmatory.'" We therefore have attempted to deal only with some of the prosodic features that appear to have a clear linguistic function and that have been investigated by means of quantitative procedures.

REFERENCES


203


Haggard, M. 1969. Pitch is voicing; manner is place. JAcS. 46.97.


Liberman, P. 1968a. Direct comparison of subglottal and esophageal pressure during speech. JAcS. 43,1157-64.
Müller, J. 1848. The physiology of the senses, voice and muscular motion with the mental facilities, W. Baly, translator. London: Taylor, Walton and Maberly.


