XVII. SPEECH COMMUNICATION

Academic and Research Staff

Prof. Kenneth N. Stevens	Dr. Allan R. Kessler	Dr. Jaqueline Vaissière‡
Prof. Morris Halle	Dr. Dennis H. Klatt	Dr. Katherine Williams
Dr. Sheila Blumstein*	Dr. Paula Menyuk†	Dr. Catherine D. Wolf**
Dr. William L. Henke	Dr. Joseph S. Perkell	Mary M. Klatt
Dr. A. W. F. Huggins	Dr. Raymond A. Stefanski	Lise Menn

Graduate Students

Thomas Baer	Bertrand Delgutte	Ursula G. Goldstein
Marcia A. Bush	William F. Ganong II	Shinji Maeda
William E. Cooper		Victor W. Zue

RESEARCH OBJECTIVES AND SUMMARY OF RESEARCH

1. Studies of Speech Production and Perception

National Institutes of Health (Grant 5 ROl NS04332-12)

Morris Halle, Dennis H. Klatt, William L. Henke, Kenneth N. Stevens

a. Larynx Mechanisms and Fundamental-Frequency Variations in Speech

Our research on laryngeal mechanisms involves physiological studies of larynx vibration, acoustical studies of fundamental-frequency variations in speech, and theoretical formulations of glottal mechanisms. An investigation of the vibration pattern of excised dog larynxes has led to a detailed description of the wave motion that occurs on the surfaces of the vocal cords, and to estimates of the aerodynamic forces that act on the vocal cords in order to maintain vibration. The nature of the vibration pattern of the vocal cords for various sounds in normal speech is being studied indirectly by observing the detail of the signal from a small accelerometer attached to the external neck surface. Careful interpretation of this waveform provides information on the opening and closing times of the glottis and on the subglottal resonances.

In studies of the fundamental-frequency contours of speech we are examining the form of these contours in sentences in English and French. In both languages the contours can be segmented into sense groups or phonetic groups that in turn are composed of elements consisting of simple rises, falls, plateaus, or prominences in the contour. Attempts are being made to establish the physiological correlates of these elements, in terms of gestures involving the intrinsic and extrinsic laryngeal muscles. Some individual differences have been observed in the way these components of a contour are actualized and in the manner in which a speaker segments an utterance into groups.

b. Timing Studies

One aspect of our research on timing in speech is aimed at the formulation of a model that will account for the timing of segments in sentences in English. Some of the

---

*Assistant Professor, Department of Linguistics, Brown University.

†Professor of Special Education, Boston University.

‡Instructor of Phonetics, Department of French, Wellesley College.

**Assistant Professor, Department of Psychology, Wellesley College.

experimental data that contribute to this model include measurements of segmental durations in carefully designed lists of sentences. These data are being supplemented by measurements on a large number of segment durations of longer samples of relatively unrestricted spoken text. The approach is to adjust the parameters of the model until it accounts for the main durational effects observed in the various syntactic and phonetic environments in the spoken material, at least for one speaker.

The current version of the model incorporates the following framework and specific rules. An inherent duration is assigned to each phonetic segment of a sentence that is to be produced. This is the longest duration that one would expect for that segment type in a nonemphatic environment. All rules are segment-shortening rules. An incompressibility constraint is included to simulate the fact that segments shortened by one rule resist the shortening influence of other rules. Specific segment-duration adjustment rules account for the following effects: Phrase-final syllables are longer in duration; word-final syllables are slightly longer; stressed and prestressed segments are longer; features of the postvocalic consonant influence vowel duration; and durational reorganization occurs in consonant clusters.

In a series of experiments we are examining the intelligibility of speech that is segmented into short intervals of approximately 60 ms, separated by silent intervals of various duration. These experiments provide insight into the short-term memory processes involved in the decoding of speech.

c. Speech Production and Perception at the Segmental Level

As part of our continuing effort to formulate a theory that will account for the various places of articulation that are found for consonants in various languages, we have been examining the production and perception of retroflex stop consonants. The production is studied through acoustic analysis of these consonants produced by native speakers of languages that contrast retroflex and nonretroflex consonants. The effectiveness of the various acoustic attributes as cues for the perception of retroflex consonants (as contrasted with velar and with nonretroflex coronal consonants) is examined through experiments in which consonant-vowel syllables with various attributes are synthesized and presented to listeners for evaluation. The findings appear to demonstrate the existence of "quantal" acoustic attributes associated with retroflex place of articulation. Present experimental studies are devoted to examining the discrimination of items in a series of consonant-vowel syllables that encompass the retroflex category by speakers of English and of Hindi.

A similar set of experiments is planned to investigate the acoustic attributes and their perceptional relevance for the voice-voiceless distinction in languages such as Spanish and Hindi where this distinction occurs without being accompanied by a difference in aspiration (as in English).

Experiments are in progress to determine the manner in which consonants characterized by rapid spectrum change are analyzed and identified, particularly the potential role of property detectors in this process. These experiments include examination of the role of the initial energy burst as a cue for place of articulation for stop consonants, studies of consonant perception in the presence of interfering adjacent speechlike stimuli, and studies in which the technique of selective adaptation is used.

d. Speech Synthesis by Rule

In recent years we have developed a set of rules for synthesizing sentences from a specification of the input phonetic string. We are now testing these rules by synthesizing a long sample of speech and comparing in detail spectrograms of the synthesized speech with spectrograms of the same passage spoken by one individual. This process has revealed inadequacies in our synthesis rules, and is leading to a reformulation and a new implementation of the rules in a somewhat different format.

e. Physiological Modeling

As part of our research aimed at understanding the control of speech production, a dynamic model of the tongue has been simulated. This model represents the tongue volume by a series of points located within the tongue body, as well as on its surface. These points are held together and activated by elements that simulate connective tissue and muscle. In preliminary studies of the behavior of the model we have examined the effects of various muscles on tongue shape, and have simulated the production of several vowels. Research plans for the forthcoming year are expected to include several physiological studies, some of which are aimed at gaining a better understanding of the relation between electromyographic records and articulatory positions and movements. The purpose of these studies is to provide a basis for utilizing electromyographic data from real speech to infer excitation patterns for muscles in the model.

f. Sound Production in Birds

With support from a National Institutes of Health Fellowship, sound production in the syrinx of various species of birds is being studied. The techniques that are employed are similar to those used in studies of larynx behavior in humans. Included are anatomical studies, detailed acoustical analyses of the sounds, studies of the vibration patterns in excised syringes by stroboscopic techniques, analysis of sounds produced by birds breathing a helium-oxygen mixture, and attempts to model aerodynamic, mechanical, and acoustical processes. These studies are helping to indicate the role of mechanical vibratory components and acoustic resonances in sound production.

2. Acoustic Studies of Speech Sounds: Invariant Attributes and Speaker Differences

U. S. Navy Office of Naval Research (Contract N00014-67-A-0204-0069)

William L. Henke, Kenneth N. Stevens

In two projects we are examining in quantitative terms the acoustic properties of selected speech sounds produced by several speakers in a controlled phonetic environment. In both of these projects the acoustic attributes are extracted with the aid of linear prediction techniques. A third project, which is nearing completion, involves measurement of formant trajectories for a number of diphthongs, diphthongized vowels, and r-colored sounds produced by 10 different speakers. The data have shown surprisingly large individual differences in target formant frequencies for some of the sounds, which suggests that there are small "dialectal" differences from one speaker to another.

We are examining the detailed acoustic properties of prestressed consonants and consonant clusters produced by several speakers under a highly controlled phonetic environment. This work involves collection of a large corpus of data, and storing, processing, and observation of these data with the aid of computer facilities at M. I. T. Lincoln Laboratory. The objective of this research is to gain a more detailed understanding of the relationship between the acoustic realization of the speech sounds and their underlying phonetic features. Of particular interest are measurements with fine time resolution but gross frequency resolution of the temporal and the spectral characteristics occurring in the frication, aspiration, and voicing phases of stop consonants in various vowel environments.

JSEP

3. Computer-Aided Signal Processing: Higher Level Dialogues and Systems for Signal Processing

Joint Services Electronics Program (Contract DAAB07-74-C-0630)

William L. Henke

The objective of this research is to design higher level "dialogue type" languages for signal processing, and to design computer-aided systems to support such interactive signal analysis and synthesis. The state of the art in the design of specific (lower level) signal-processing algorithms and the design and implementation of related hardware has recently advanced rapidly, but the benefits of this progress have not been readily available to problem-oriented users who are experts in their problem area rather than in the implementational intricacies of signal processing. Such users need a system that allows them to integrate selected primitive operations into larger processes quickly and easily (that is, online), and then to apply these processes to their data. They should be able to adjust process parameters and to select displays and other outputs of a variety of formats that might emphasize features of interest for their particular problems. Our goal, then, is to evolve a dialogue and system which will allow problem-oriented users to bridge the "linguistic gap" between themselves and the signal processors.

During the past six months representation and dialogue techniques for "synchronous sequential" types of processes have been considerably improved and augmented, specifically in regard to block-diagram editing techniques, the treatment of multiple input primitives, the inclusion and design of transversal filters, and the development of block-diagram "compilation" techniques needed to support more general topological structures. The "Fourier component" of the system has been augmented to support greater flexibility and power in definition and application of spectral and correlational types of operation. The dialogue and system are described in an unpublished document (which is being updated continually), entitled "MITSYN – An Interactive Dialogue and Language for Time Signal Processing."[1]

For such signal-analysis systems to be economically viable for general field usage, they must be capable of implementation as systems no larger than those of the minicomputer class. The possible scope of the dialogue, however, is such that the implementation of the system requires large amounts of software. Thus the only reasonable approach to the implementation of such systems is to write them in higher level programming languages featuring (i) a clean and powerful language for programmer productivity, (ii) run-time efficiency for use on minicomputers, (iii) machine independence for transportability, and (iv) availability. Unfortunately, such programming languages do not seem to exist. Thus a part of our effort has been devoted to designing programming languages and implementing translators. Two such programming languages have now been sufficiently developed to support almost all of the programming for the signal-processing activities. One is called "Minicomputer BCPL" and the other "Structured FORTRAN." These languages are described in unpublished programmers' manuals.

References

1. W. L. Henke, TM-1, Research Laboratory of Electronics, M. I. T., February 1975 (unpublished).

JSEP

A. FURTHER NOTE ON FRENCH PROSODY

National Institutes of Health (Grant 5 RO1 NS04332-12)

Jacqueline Vaissière

1. Introduction

This report continues our analysis[1] of fundamental frequency (F_o) contours in French. The previous study has been extended by giving a modified version of the paragraph (Reading Material) that appeared in the appendix of the previous report and by adding a list of new isolated sentences to the speech that has been analyzed. The new corpus serves as an appendix to this report. Three more speakers have been used as subjects, and two of our previous speakers also read the new corpus.

The previous report[1] made the following points.

1. The speaker decomposes the sentences by inserting pauses into sense groups (each sense group being composed of closely related words), and the number of sense groups in a sentence tends to decrease as the rate of elocution increases.

2. The intonation patterns of words in final position of a sense group are determined by the position of that sense group in the sentence (intonation of the words, a falling or a rising, depends on whether the sense group is in final position in the sentence). We noted that words that are not in final position in the sense group are pronounced with certain types of intonation patterns depending on the speaker's judgment of the closeness of the relation between the successive words.

Our analysis of the actual shape of the F_o contours for each word within a sense group leads us to conclude that F_o contours of any word can be schematized visually by one of the four patterns represented in Fig. XVII-1 which represents F_o contours found for words containing three or more syllables. The patterns for shorter words (monosyllabic and disyllabic) often degenerate into simpler contours.

We shall now describe the four patterns in detail and then discuss the factors that influence the speaker to choose a particular pattern when he pronounces a word embedded in a sentence.

2. Contour Patterns

The schematized F_o contours of the words are described in terms of certain attributes, such as Rising, Lowering or Peak, which were used by Shinji Maeda in his study of characterization of F_o contours for American English.[2]

a. Pattern P4 (Ri+F)

The pattern P4, shown in Fig. XVII-1a, is characterized by an initial rise (Ri) followed by a falling contour (F). This pattern was found for the final word at the end

of a sentence, or for words in nonfinal position in a sense group. The falling contour always reaches a lower F_o value in sentence final position than in nonfinal position

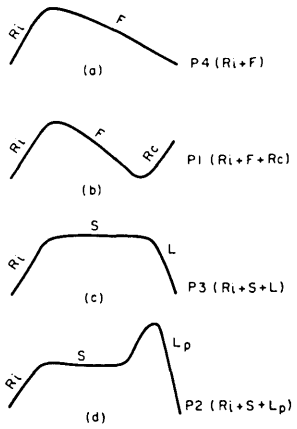


Fig. XVII-1.

Schematized F_o contours for the long words embedded in a sentence. The six attributes are as follows:

Ri: Initial Rise
Rc: Continuation Rise
L: Lowering
Lp: Lowering associated with Peak
S: Sustained
F: Fall

within a sentence. The initial rise (Ri) occurs at the onset of the word (more details on Ri will be given later). The falling F immediately follows the rise Ri, and continues until the end of the word. The extension of falling to the onset of the following word seems to depend on various factors such as the rate of elocution and the habits of individual speakers.

b. Pattern P1 (Ri+F+Rc)

The pattern P1, shown in Fig. XVII-1b, was found at the end of a sense group which is followed by a pause; only one of the nine speakers used P1 for words inside a sense group, followed by a very short pause. Pattern P1 is primarily characterized by a rising contour at the end of the word. The F_o contour is raised at the beginning of the word (Ri) and then falls (F) as observed in the previous pattern P4, but is terminated by the continuation rise Rc (following the terminology of Delattre[3]) during the last-sounded syllable of the word. In disyllabic words, the pattern P1 can be realized fully as Ri+F+Rc, but it often degenerates into a simpler contour so that the important attribute for this pattern, Rc, is still conserved. Therefore the schematic pattern for P1 in short words may be described as

$$P1 = \left\{ \begin{matrix} Ri \\ \emptyset \end{matrix} \right\} + \left\{ \begin{matrix} S \\ \emptyset \\ F \end{matrix} \right\} + Rc$$

where one of the elements in { } has to be chosen, and $\emptyset$ indicates a deletion.

c. Pattern 3 (Ri+S+L) and Pattern 2 (Ri+S+Lp)

The patterns P3 and P2 (Fig. XVII-1c and 1d) were found for the nonfinal words in the sense groups. They are characterized by a sustained F_o contour (S) which indicates a plateau during the intermediate syllables of the word. During the plateau, the F_o values may fall gradually along the syllables or stay approximately level (depending on the speaker). In both patterns P3 and P2 the F_o rises at the onset of the word, and is then sustained. For pattern P3, the F_o then falls rapidly at the end of the word (during the last phoneme or during the whole last syllable) and the lowering contour (L) is connected to that of the following word. In pattern P2 the initial rise Ri and the sustaining S are similar to Ri and S in P3, but the lowering L in the last syllable is preceded by a sharp rise. The successive rise and fall during the last syllable of the word indicate a peak (p) in the F_o contour. The rise occurs at the onset of the last syllable, and involves generally only the first phoneme of that syllable.

3. Correspondence between the Attributes and the Phonemes

Each attribute corresponds to a certain portion (sequence of successive phonemes) of the word. We have already described the parts of the word corresponding to the attributes L, Lp, and Rc. We shall now describe the portion taken by Ri (the sequence of phonemes taken by the attributes F and S are automatically determined if the portions taken by the other four attributes are known). The portion corresponding to the initial rise Ri depends on the phonemic category at the onset of the word. Three separate cases must be considered: in case 1 the initial phoneme is a consonant other than /l/, /m/ or /n/; in case 2 it is one of the consonants /l/, /m/ or /n/; in case 3 the initial phoneme is a vowel. In these three cases we define the portion of the rise Ri as the segment between the end of the preceding word and the phoneme in which the rising Ri is connected to F (in P1 and P4) or to S (in P3 and P2). This definition can be adopted for the rise Ri that occurs during the unvoiced consonants in case 1, where the F_o contour is interrupted; for the initial voiced stops and fricatives and the consonant /r/, the F_o contour often contains a valley (micromelody), but we recognize the valley as a segmental effect, and hence ignore it.

In case 1, where the initial phoneme is a consonant other than /l/, /m/, /n/, Ri occurs during that consonant, and occasionally (for some speakers) the rise Ri is extended to the following vowel. See, for example, the F_o contour of the word "la compatibilité" in Fig. XVII-2a, the word "la parasitologie" in Fig. XVII-3, or the word "la confédération" in Fig. XVII-4a. (In the three examples rise Ri occurs during the

voiceless consonant; other examples were given in our last report.[4]) In case 2, where the initial phoneme is the consonant /l/, /m/ or /n/, the rise Ri occurs at least during that initial consonant and the following vowel. See, for example, the word "manifestations" in Fig. XVII-4b [and also the words "météo" and "Massachusetts" in Figs. XVI-22 and XVI-26 (Quarterly Progress Report No. 114, pages 214 and 219)]. Often the rise Ri is extended to the following sonorant segment, if any. See, for example, the F_o contour of the word "manifestations" in Fig. XVII-5 pronounced by a different speaker from the one of Fig. XVII-4. In that case the rise is extended to the two first syllables. When the word begins with a vowel (case 3), the rise Ri occurs at least during the first sonorant segment of the word (we call a sonorant segment a sequence of phonemes composed of a vowel and the consonants /l/, /m/, /n/). Compare, for example, the F_o contours in the words "la compatibilité" and "l'incompatibilité" in Fig. XVII-2;

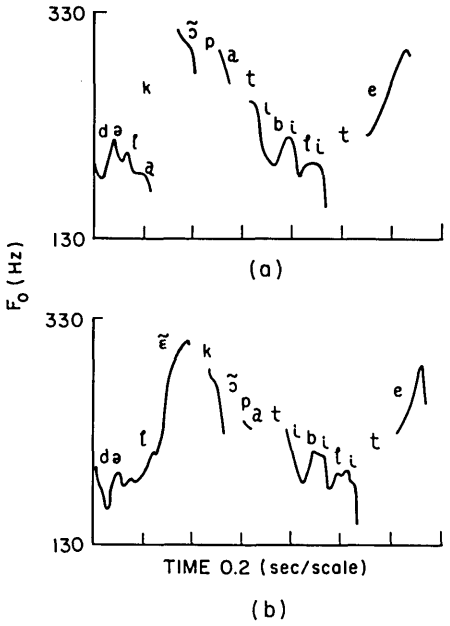
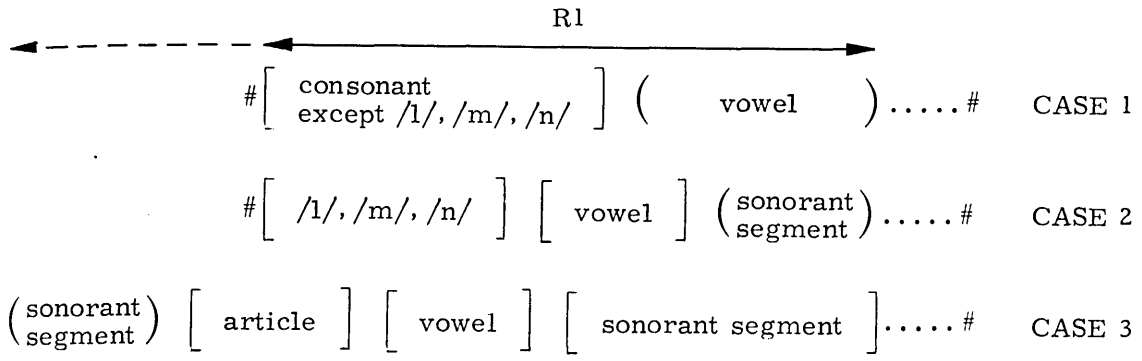


Fig. XVII-2.

F_o contours of the words "de la compatibilité" and "de l'incompatibilité" embedded in a sentence.

in the second word the portion of Ri includes the first syllable. The F_o contour of the word "l'anesthésiologie" in Fig. XVII-3 shows the extension of Ri to the second (sonorant) syllable of the word. In case 3 the rise Ri may also be extended to the sonorant segment of the preceding function word, particularly at the beginning of a sense group. See, for example, the extension of the rise Ri to the preceding article "une" in the sense group "est une institution mixte" in Fig. XVII-7a and XVII-7b.

We may summarize the preceding results by the following diagram: brackets indicate the portion taken by Ri, parentheses indicate an optional extension of Ri, and (#) indicates the lexical word boundaries.



Note that the portion of Ri is closely related to the duration of the phonemes. For words located in a similar position in a sentence and pronounced with one of the four patterns (P1, P2, P3 or P4), the angle of the slope of F_o rising during Ri tends to stay constant, independent of the identity of the phonemes, and the duration of the rising portion tends to be kept fixed, regardless of the number of phonemes underlying the rising. Figure XVII-3 shows the superimposed four F_o contours of the words "la parasitologie," "la radiologie," "l'anesthésiologie," and "la cardiologie" spoken by one speaker (the four words compose the subject phrase in a sentence). We note that the overall contours

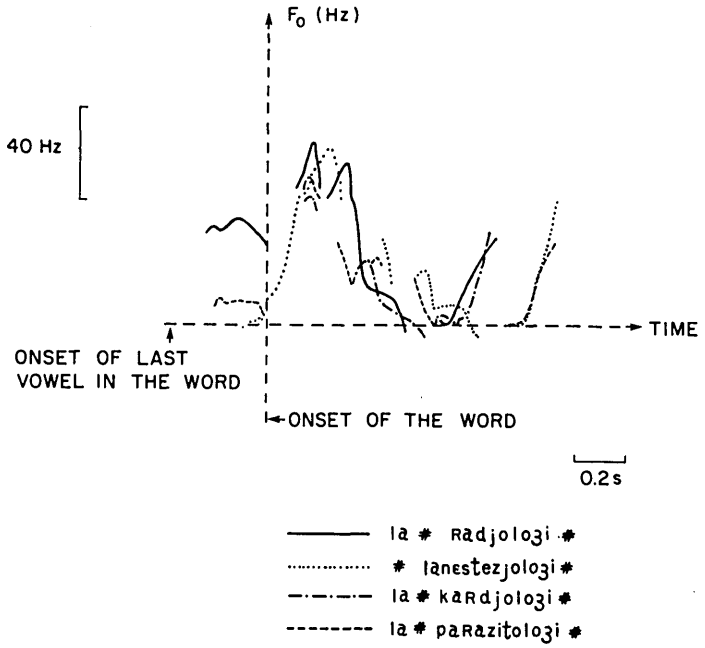


Fig. XVII-3. Superimposed F_o contours of the four words "la radiologie," "l'anesthésiologie," "la cardiologie," and "la parasitologie." The F_o contours of each word are shifted along the frequency scale to line up the F_o values at the onset of the last vowel (/i/) in the words.

for the four words are quite similar. We also observe that if the distribution of the portions for the attributes Ri and F differs from one word to another, Ri depends essentially on the phonemic context, Rc depends on the habit of the speaker (in Fig. XVII-3, the speaker raises F_o only during the last vowel), and the fall F takes the remaining portion. In Fig. XVII-4, F takes two syllables in the words radiologie and cardiologie, three syllables in anesthésiologie and four syllables in parasitologie.

4. Nuclear Stress Rule within the Sense Group in French

What factors influence the speaker when he chooses one of the three patterns P2, P3, or P4 for the words inside the sense groups? The most common case is sense groups composed of three lexical words. Suppose that Px, Py, and Pz represent the patterns attributed to the first, second, and last lexical words. In nonfinal sense groups we have found all possible combinations for the pattern of the first word, Px, and the pattern for the second word, Py. The combinations of the pattern for the first and second words can be separated into two cases: in case 1 the speakers repeat the same pattern (in that case, Px = Py), and in case 2 they choose two different patterns for Px and for Py (i.e., Px ≠ Py).

In the first case, the speakers generally choose their most frequent pattern for Px and Py, either P2, P3, or P4. For example, the subject noun phrase "L'Institut de Technologie du Massachusetts..." has been pronounced with the following descriptions depending on the speakers: P4P4P1 by two speakers, P3P3P1 by two speakers [see an example in Fig. XVI-26d (Quarterly Progress Report No. 114, page 219)], and P2P2P1 by three speakers (see Fig. XVI-26a in Quarterly Progress Report No. 114, page 219). In this case the distribution of the patterns does not give information about the internal syntactic structure of the sense group. We call a combination where Px = Py a prosodic parallel structure. Prosodic parallel structure has been found frequently for sense groups representing a single meaning (such as an institution or an organization) or a noun phrase [such as the noun phrase subject "un retour offensif de l'hiver" represented in Fig. XVI-23 (Quarterly Progress Report No. 114, page 216, with the description P4P4P1)].

In the second case, however, the speaker used two different patterns for Px and Py (Px ≠ Py). Figure XVII-4 illustrates the complete F_o pattern for the clause "La confédération générale du travail a organisé des manifestations importantes..."; the clause is divided by a pause into two sense groups, the first one equivalent to the subject noun phrase (represented in Fig. XVII-4a) and the second to the predicate (represented in Fig. XVII-4b). The first sense group may be described by the sequence of patterns P4P2P1, and the second by the sequence P2P3P1. The noun-phrase subject has been spoken generally with a parallel prosodic structure (by 6 speakers) and also with the

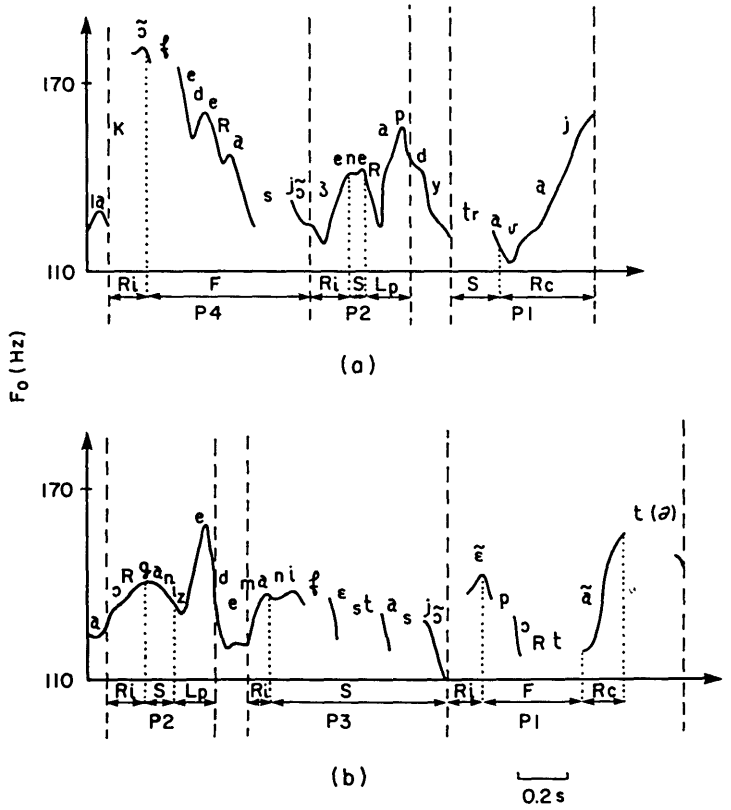


Fig. XVII-4. F_0 contour of the clause "La confédération générale du travail a organisé des manifestations importantes . . ." spoken by speaker BD.

description P3P2P1. We found one parallel prosodic (P2P2P1) for the predicate for one speaker; the other speakers used either P2P3P1 (Fig. XVII-4b) or P2P4P1 or P3P4P1 (Fig. XVII-5). In case of a syntactic left-branched structure (such as the noun-phrase subject in Fig. XVII-4a) the speakers realized it with a parallel prosodic structure (Px = Py), or with a left-branched prosodic structure, such as P4P3P1, P4P2P1 or P3P2P1. This may be described as PxPyP1 where (x⟩y)). Similarly, the speakers realized a syntactic right-branched structure such as the predicate represented in Figs. XVII-4 and XVII-5 with a parallel structure or with a right-branched prosodic structure such as P3P4P1, P2P4P1 or P2P3P1. This may be described as PxPyP1 where (x⟨y)). Figure XVII-6a shows the organization of the pattern in nonfinal sense groups composed of 3 lexical words.

There is an interesting analogy between the stress pattern for a phrase derived by Chomsky's Nuclear Stress Rule[5] in English and the sequence of the patterns PxPyPz that we have described. In the case of a left-branched structure, the Nuclear Stress Rule for English assigns the degrees of prominence 3, 2, and 1 to the lexically stressed

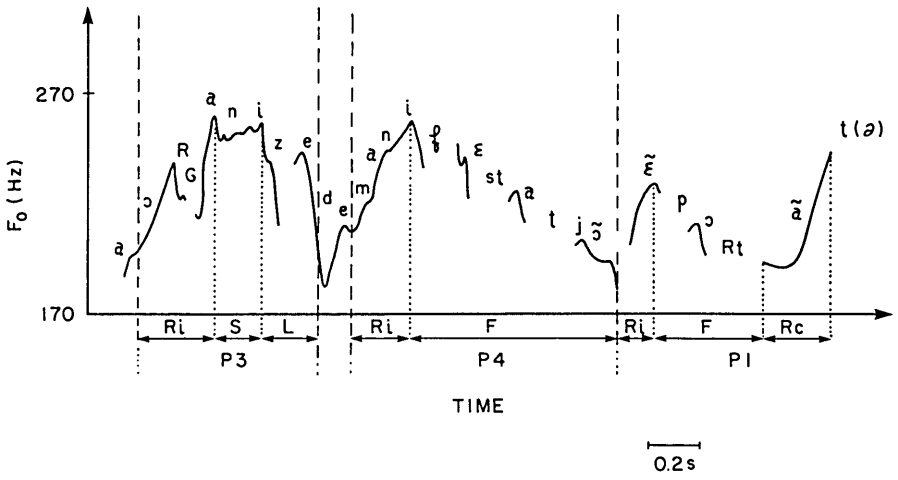


Fig. XVII-5. F_0 contour of the predicate "a organisé des manifestations importantes ..." spoken by speaker MP.

SYNTAX	PROSODY	
	Nonfinal Sense Group (Z=1)	Final Sense Group (Z=4)
Left-branched Structure	Left-branched or Parallel Prosodic Structure	
P_x P_y P_z (Chomsky-Halle Rule 321)	$x \geq y > z$ x and y = 2,3 or 4	$z > x \geq y$ y = 2 or 1 (insertion of pause) x = 2,3 or 4
Right-branched Structure	Right-branched or Parallel Prosodic Structure	
P_x P_y P_z (Chomsky-Halle Rule 231)	$y \geq x > z$ x and y = 2,3 or 4	$x \leq y < z$ y = 2 x = 2 or 1 (insertion of pause)

Fig. XVII-6. Organization of the F_0 patterns in nonfinal and final sense groups composed of three lexical words.

syllables of the first, second, and third words, respectively; it assigns the degree of prominence 2, 3, and 1 in case of a right-branched structure. In all cases, number 1 designates a higher degree of stress. A rule similar to the Nuclear Stress Rule would also be valid for French, but it would not assign degrees of prominence: it would assign certain F_0 patterns or a possibility of certain combinations of patterns to the words. The more stressed syllables in English correspond to the more rising syllables in French: P1 where the last syllable has a rising intonation, P2 where the last syllable has consecutively a rising and a falling contour, P3 where the last syllable is sustaining and falling, and P4 where the whole word has a falling intonation.

When the sense group is sentence-final, there are less attested possible combinations (see Fig. XVII-6). First, the last word has a falling intonation (Pz = P4). And, as we have found previously,[1] the final fall starts from the last syllable of the penultimate lexical word, which has pattern P2. There are a few exceptions where the penultimate word has pattern P3, and there is one case of pattern P4. Consequently, speakers only exercise choice in the pattern of the first word, which can be P4, P3, or P2. In case of a left-branched structure, the speaker inserts generally a small pause after the second word and P2 becomes P1, or pronounces the sense group with a parallel prosodic structure, P2P2P4. In case of a right-branched structure, the most attested combination of patterns is also P2P2P4. The opposition between right- and left-branched structure is neutralized in most cases, because of the constraints on Py and Pz. To disambiguate phrases like "un marchand de tapis persans" and "un marchand de tapis persan" (in the first phrase, the carpets are Persian, and in the second, the seller is Persian), most speakers inserted a pause in the second case and the most attested distributions of patterns for the two cases are P2P2P4 and P2P1P4. The first case when heard in isolation is still ambiguous. When the sense group is composed of only two words, the first word may have one of the three patterns, P2, P3, or P4 in a nonfinal sense group, but generally only P2 in a final sense group. Figure XVII-7 illustrates the sense group "est une institution mixte" pronounced by three

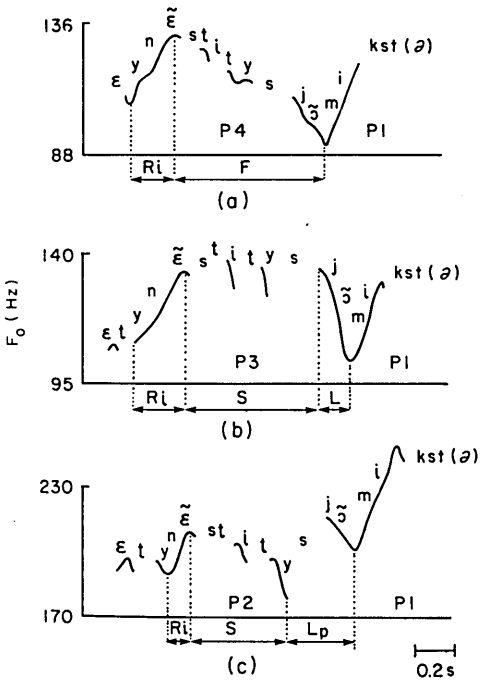


Fig. XVII-7.

F_0 contour of the words "est une institution mixte" spoken by three speakers. Each speaker uses a different pattern (P4, P3 or P2) for the word "institution."

different speakers, each of whom chooses a different pattern for the first word: P4 for the first speaker, P3 for the second, and P2 for the third. When they speak rapidly P4 is generally preferred by speakers to the two other patterns, probably because it requires less effort for this particular example.

5. Conclusion

A phonetic unit such as the sense group that we have described is found not only in French but also in English. Shinji Maeda[2] claimed that American-English sentences are demarcated into smaller units, which he called "phonetic groups." He composed a certain type of F_o pattern corresponding to each group. The contents of the sense group and the phonetic group in French and in English are quite similar and correspond to linguistics constituents such as noun phrases or predicates. The patterns differ radically, however, in French and in English. In English, the F_o pattern for a phonetic group indicates roughly a "hat pattern." The contour is raised at the first primary syllable in the first content word and is lowered at the end of the primary stressed syllable in the final word, so that the lexical stresses are well represented by a rapid rising and a rapid falling of F_o values. In French, on the other hand, the sense groups are distinguished by certain combinations of F_o patterns for each word inside the sense group. This difference may arise because in English sense groups there are lexical stresses constraining the F_o pattern, whereas in French such stresses are absent and hence F_o patterns are applied to words as a whole.

Appendix

Reading Material 2

1. Modified Version of the Previous Paragraph:

L'institut de technologie est une institution privée et mixte, dont les centres d'intérêts sont l'architecture, les sciences pures et les sciences de l'ingénieur. Il a apporté sa contribution aux progrès technologiques des dernières années, et il continue sa participation dans les techniques les plus récentes. La gamme de ses recherches est très étendue, et elle s'étend de l'électronique à la biologie, en passant par les sciences nucléaires, la linguistique et l'économétrie. Les étudiants peuvent suivre des cours très divers et participer à des recherches très variées, en profitant de l'association de l'institut avec les universités des environs. C'est à la fin de la seconde guerre mondiale que fut construit le laboratoire de recherche en électronique. Une centaine de professeurs, encadrant trois cent cinquante étudiants y conduisent des recherches.

2. Isolated Sentences:

The following phrases are embedded in two carrier sentences so that each phrase forms a final and a nonfinal sense group.

Un professeur de sociologie américain

Un professeur de sociologie américaine

Un marchand de tapis persan

Un marchand de tapis persans

Un spécialiste de la géographie de l'Amérique

Un professeur de géographie d'Amérique.

References

1. Jacqueline Vaissière, "On French Prosody," Quarterly Progress Report No. 114, Research Laboratory of Electronics, M. I. T., July 15, 1974, pp. 212-223.
2. Shinji Maeda, "A Characterization of Fundamental Frequency Contours of Speech," Quarterly Progress Report No. 114, Research Laboratory of Electronics, M. I. T., July 15, 1974, pp. 193-211.
3. P. Delattre, "La leçon d'intonation de Simone de Beauvoir," French Rev. 35, 59-67 (1961).
4. Quarterly Progress Report No. 114, pp. 212-223, the words "consonnes" and "finales" in Figs. XVI-24 and XVI-25, and the word "technologie" in Fig. XVI-26.
5. N. Chomsky and M. Halle, The Sound Pattern of English (Harper and Row Publishers, Inc., New York, 1968).

B. ELECTROMYOGRAPHIC STUDY OF INTONATIONAL ATTRIBUTES

National Institutes of Health (Grant 5 RO1 NS04332-12)

U. S. Navy Office of Naval Research (Contract ONR N00014-67-A-0204-0069)

Shinji Maeda

In a previous report[1] it was suggested that F_o contours of American English sentences are characterized by a limited number of intonational attributes: a rising, R, and a lowering, L, which form a piecewise-linear trapezoidal pattern ("hat pattern"), a peak, P, which is often associated with the rising R, and a rising R1 that is a rise on a plateau of the hat pattern. The words in a sentence are grouped into larger units, which we call phonetic groups (PG's), by forming a hat pattern for each PG. Inside PG the fundamental frequency (F_o) is raised during the primary stressed syllable in the first lexical word and is lowered from the end of a stressed syllable in the last word. A syllable between the two stressed syllables can receive R1, depending on the structure of PG. The manner of grouping the words into PG's depends not only on the structure

of the sentence, but also on some other factor, perhaps an economical principle of physiology. It was emphasized that experimental study of the physiology of laryngeal control was needed to obtain a deeper understanding of F_0 control during speech.

In this report we shall describe the results of an electromyographic (EMG) experiment on the intrinsic and extrinsic laryngeal muscles. The experiment was performed at Haskins Laboratories, for one speaker (KNS) of the three who were the subjects of the previous study.

1. Experimental Procedure

A corpus composed of 27 isolated sentences and a text were used for this experiment. Data from 23 sentences of these 27 sentences are in this report. EMG signals were obtained from several intrinsic and extrinsic laryngeal muscles: cricothyroid (CT), vocalis (VOC), lateral cricoarytenoid (LCA), sternohyoid (SH), sternothyroid (ST), and mylohyoid (MH). Bipolar hooked-wire electrodes[2] were inserted into each of these muscles to obtain the EMG signals. Two sets of 27 sentences written on cards were randomized separately, and the subject read each sentence as the card was shown. The two sets of cards were presented alternately 8 times so that each sentence was read 16 times. The raw EMG data signals then were processed by the Haskins Laboratories EMG Data System.[3,4] The curves representing the EMG activities were obtained by integrating the raw EMG data signals for each sentence over a 10-ms time window and then averaging over from 12 to 16 repetitions of the sentence. The same window length was used for each of the six muscles.

F_0 contours were obtained by using our M. I. T. computer facility, as described in the previous report.[1]

2. Results

a. Relationship between the Attributes of the F_0 Contour and the Activities of the CT and the SH Muscles

Comparative study of the F_0 contours and the corresponding EMG data for each muscle showed that CT activity and SH activity are relevant to the control of F_0, and thus to the attributes rising R, peak P, lowering L, and rising Rl. The ST activities are quite similar to the SH activities pointed out by Atkinson[5] and Collier,[6] although the two activities differ considerably from each other for certain phonemes; for example, during /l/ ST was much more active than SH in terms of EMG for that experiment.

In Fig. XVII-8, the F_0 contour and corresponding EMG activities from CT and SH are shown for sentence S6 "The farmer raises sheep." The idealized F_0 patterns

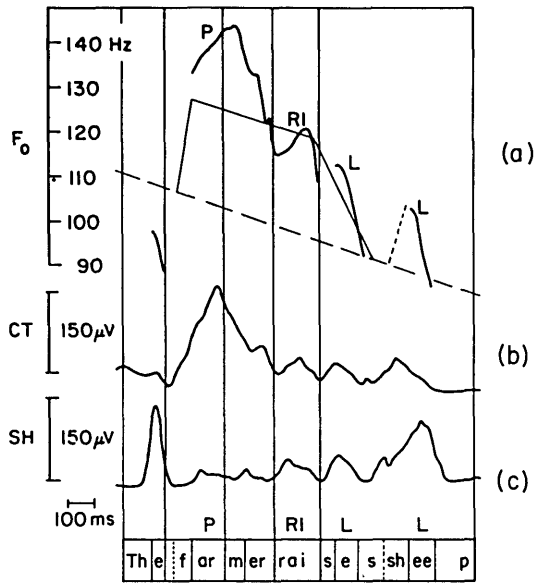


Fig. XVII-8. (a) F_0 contour associated with its schematized pattern. (b) The corresponding averaged EMG activity for the CT muscle, and (c) for the SH muscle.

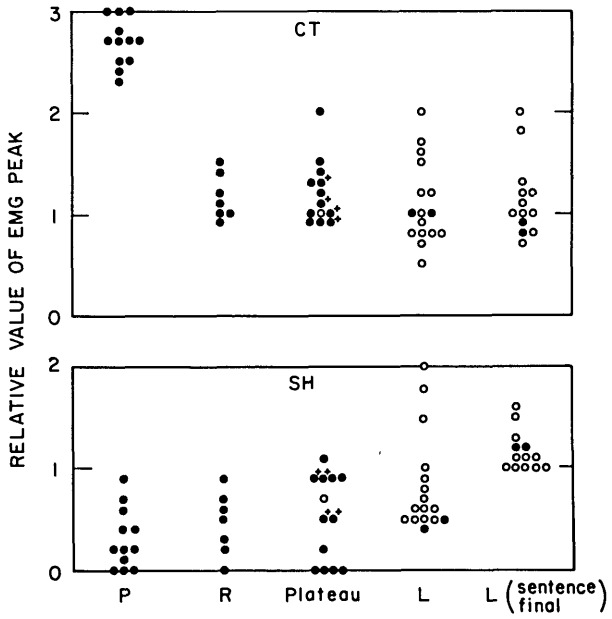


Fig. XVII-9. Relative values of peak heights of the EMG activities for CT and SH muscles, within the syllable classified by its attribute determined from the F_0 contour. Closed circles represent the SH peak preceding the corresponding CT peak, and open circles the reverse temporal relationship.

marked with the attributes are superimposed on the F_o contour. The F_o contour is computed from one of 12-16 repetitions of the sentence, and it is not an averaged contour. To compensate for the time delay of the effect of the muscle activities upon F_o values, both CT and SH curves are shifted +80 ms along the time axis. It is observed that the CT trace in Fig. XVII-8b indicates a peak for every stressed syllable in the sentence. In particular, CT is very active during the syllable with the attribute peak P. The SH activities seem to be related to the lowering of F_o. The F_o contour for "the" at the beginning of the sentence indicates very low-frequency values; a large peak can be seen in SH activities in this syllable, as shown in Fig. XVII-8c. In the syllable that receives attribute R1, as in the stressed syllable in "raises," both CT and SH are active, and it is observed that the peak of the SH activities precedes that of the CT. On the other hand, when the syllable receives attribute lowering L, as in the word "sheep," the temporal relation is reversed: the CT peak precedes the SH peak (this temporal relationship was also observed by Atkinson[5] and Collier[6]).

The analysis of the EMG curves has suggested that the relative height and the temporal relationship of CT and SH peaks within a syllable seem to be meaningful measures of the muscle's activities. A summary of measurements for 13 sentences, which are listed in the appendix, is shown in Fig. XVII-9. The relative values of peak height in the CT and SH curves are plotted for each syllable classified by its attributes determined from the F_o contour. The peak heights of the syllables corresponding to the plateaus of the hat patterns are represented in the plateau column where the circles marked "+" indicate that the syllable receives the attribute R1.

There is evidence that the magnitude of the CT peak distinguishes the attribute peak P from the remaining attributes. Taking account of the temporal relationship, we can fairly well distinguish lowering L from attributes rising R and R1 (and -R1 which corresponds to syllables on the plateau that do not have R1). Most of the circles representing the peak values for lowering L are open, which indicates that the CT peak precedes the SH peak in each syllable. Slightly higher SH peaks are seen at sentence-final position, labeled L (sentence-final) in Fig. XVII-9, than in the nonfinal position. Two groups of closed circles corresponding to attribute R1 (marked "+") and to -R1 (unmarked closed circles) are not well separated from each other. Although the unmarked circles tend to cluster around the origin for the data, there are three unmarked circles far above this cluster. An examination of the EMG curves shows that the peaks in CT and SH activities during each of these three syllables are close to each other. Perhaps as a consequence of this, the F_o values do not rise or fall during the syllables. We understand that these large SH peaks may be due to the phonetic influence of the speech sound that is being produced. In fact, it has been pointed out by several authors[5-8] that the SH muscle seems to involve both F_o control and articulation for a certain class of phonemes, although each of these authors postulates different weights for the two functions

of the SH muscles. Also, syllables on the plateau are not well distinguished from those that receive the attribute R. This may suggest that some other factors are involved in F_0 control. For instance, consider the fact that the SH muscle can act to lower the hyoid bone, and consequently the larynx. In order to repeat this action within a sentence, the larynx must be raised beforehand. This series of raising and lowering actions causes an up and down movement of the larynx as shown previously.[1] The variation in laryngeal position during speech may change a state of the internal laryngeal mechanism, and the effect upon F_0 values of the CT activities may vary, depending on this state.

b. Influence of Emphasis on F_0 Patterns

All sentences that have been mentioned heretofore were pronounced in a nonemphatic mode. Therefore the question may arise about whether a sentence with emphasis is characterized by the set of attributes already introduced. To obtain some perspective on this problem, we included in the corpus some short sentences with emphasis. Four utterances of the sentence "Bill meets Steve" were used: without emphasis, with emphasis on "Bill", on "meets", and on "Steve".

The schematic analysis of the F_0 contours shows the following assignments of the attributes to each of the four utterance types:

```
      P          L
S1)  (Bill meets Steve)

      P     L    RL
S2)  (BILL meets Steve)

      R     P     L
S3)  (Bill MEETS Steve)

      R     L     PL
S4)  (Bill meets STEVE)
```

where the words in capital letters are those emphasized during speaking. It is recognized that each emphasized word receives attribute P. This situation is illustrated in Fig. XVII-10 in which peak heights of the CT and SH traces within each word are plotted separately in a fashion similar to Fig. XVII-9. A markedly large peak value in CT and a small value in SH are observed in every emphasized word. This is what we found as a common characteristic of the muscle activities for attribute peak P in the previous section.

Emphasis is not only characterized by attribute P but often introduces a contrast in terms of the F_0 contour of the adjacent words. Sentences S1 (without emphasis) and S2 (with emphasis on "Bill") are distinguished by attribute L assigned on "meets" in S2. Correspondingly, in the word "meets" the largest activity in SH and a small CT

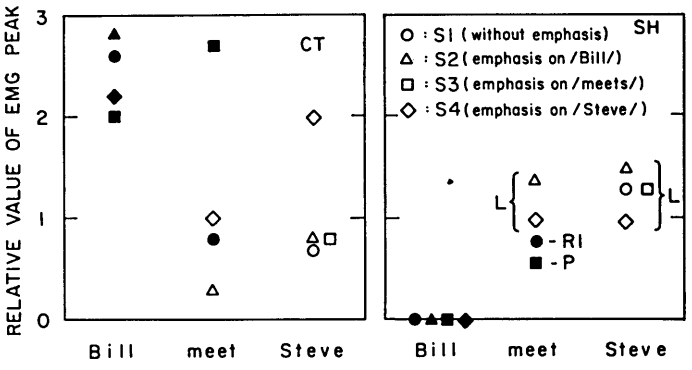


Fig. XVII-10. Relative values of peak heights of the EMG activity for CT and SH muscles within each word in sentences S1-S4. Closed dots indicate that the SH peak precedes the corresponding CT peak within the syllable, and open dots the reverse temporal relationship. Dots marked with attributes L, R1, and P indicate that each of these syllables receives that attribute.

activity are shown in Fig. XVII-10. This speaker tends to place a strong peak at the initial word of the sentences and the contrast between S1 and S2 is achieved by placing L at the word next to the emphasized one. Perhaps, for the same reason, the speaker realizes lowering L on "meets" before the emphasized word "Steve" in S4. In S3, however, such contrast activity is not introduced. The subject simply emphasizes the word "meets" by placing attribute P.

The manner of emphasis seems to depend on the speaker, since in similar experiments carried out by Ohala[7] and by Atkinson[5] the speakers seemed to reduce the peak of the initial word. If S3 is spoken in that manner, the assignment of the attributes would be like "Bill (MEETS Steve)." (P L)

In any case, we may state that emphasis on a word in a sentence is realized by locating attribute P. Furthermore, in order to achieve a clear contrast with the adjacent words, emphasis may change the local organization of the intonation pattern in the sense that different groupings of words in the same sentence may occur, depending on the location of the emphasized word as seen in sentences S1-S4.

c. Influence of Voiced and Unvoiced Stops on F_o Contours

It was pointed out previously[1] that in a stressed syllable with attribute R containing an initial voiceless consonant, the rising F_o contour cannot be seen. Also, the maximum F_o values during the following vowel are higher after voiceless than after voiced consonants.[9] The reason for this shift in F_o is not understood in detail. To obtain some insight into this phenomenon, we investigated word pairs such as "pill" vs "bill", "till" vs "dill", and "coat" vs "goat", contrasting voiced/voiceless consonants in initial

position. The same carrier sentence was used for every word. The F_0 contour for each word is raised during the initial consonant and only the lowering contour could be observed. The results are shown in Fig. XVII-11, in which the two curves for each word are superimposed, with time alignment at the onset of voicing after the stop release. The solid lines correspond to the words with voiceless stops and the dashed lines to those with voiced stops.

For each of the word pairs, we can see systematic differences that seem to be related to the voiced/voiceless contrast. These differences occur in the F_0 contours, and in the EMG data for CT, VOC, and LCA, as shown in Fig. XVII-11. The maxima of the F_0 values that occur at the onset of the voicing are consistently higher after voiceless stops than after voiced stops. This phenomenon may be explained by interpreting the corresponding muscle activity.

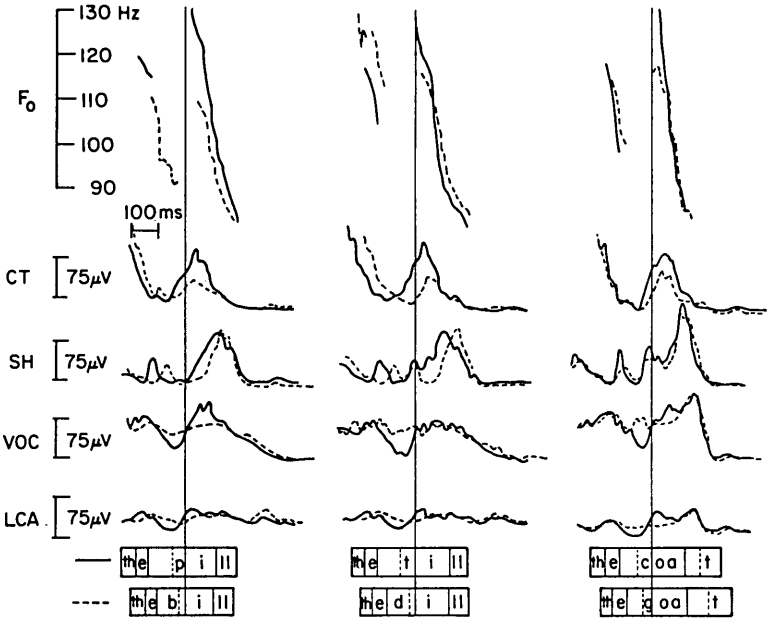


Fig. XVII-11. Comparison of F_0 contours and EMG activity curves for CT, SH, VOC, and LCA. The curves for CT and SH are shifted +80 ms along the time axis and +40 ms for VOC and LCA.

First, the peak in the CT curve after each voiceless stop is higher than that after the corresponding voiced stop. The reason why the CT curve reaches a higher value after the voiceless consonant may be the following. It should be noticed that the onset point of the CT rising for the voiceless stop always precedes that of the voiced cognate. Furthermore, each onset point of rising CT roughly corresponds to the stop release position, which is represented by the dashed line in each box

at the bottom containing the representation of the word. The longer voice-onset time in the voiceless consonant may lead to the higher CT peak during the following vowel, if it is postulated that rate of increase in CT activity and the durations of the vowel tend to be kept constant. A similar timing of the CT rise may be seen in "farmer" and "sheep" in Fig. XVII-8.

Second, the systematic variations arising from the voiced/voiceless contrast can also be found in VOC and LCA activities, as shown in Fig. XVII-11. Both VOC and LCA seem to function as adductors of the arytenoids. Deeper dips (corresponding to wider opening in the glottis) in the VOC and LCA traces are found during voiceless stops than during the corresponding voiced stops.

It is worthwhile to notice that the results for the one speaker described here seem to support a scheme of laryngeal features proposed by Halle and Stevens.[10] Hirose and Gay,[11,12] on the other hand, suggested that there was no such systematic difference in the muscle activities because of the voiced/voiceless contrast. We are not in a position to state whether these discrepancies are due to the interspeaker differences or a consequence of different experimental conditions. We strongly feel that further experimental and theoretical studies of the laryngeal mechanism are necessary to obtain more solid knowledge concerning F_0 control.

I would like to thank Dr. T. Ushijima, Dr. F. S. Cooper, and Dr. K. Harris, of Haskins Laboratories, for their kindness and help in the EMG experiment. The EMG facilities at Haskins Laboratories are supported by a grant from the National Institute of Dental Research.

Appendix

A list of the sentences from S5 to S17, marked with the attributes which were determined from the F_0 contours. The sentences from S1 to S4 are listed in the text.

```
          P      L      L
S5)   (Ken raises) (sheep).

          L   P        R1 L      L
S6)   The (farmer raises) (sheep).

          P L                    L
S7)   (All farmers) raise (sheep).

                P L                    L
S8)   Almost (all farmers) raise (sheep).

          L   P        R1 L     R        L
S9)   The (farmer raises) (yellow sheep).

          P L                  R        L
S10)  (All farmers) raise (yellow sheep).

                P L                  R        L
S11)  Almost (all farmers) raise (yellow sheep).
```

P R1 L R L
S12) (Ken raises) the (great yellow sheep).

P R1 L R L
S13) (Ken raises) the (light yellow sheep).

PL R L L
S14) In the (house), (Bill drinks a) (beer).

P L L L
S15) (Bill drinks a) (beer) in the (house).

P L R L L
S16) (Bill drinks a) (beer in the) (box).

P L L L
S17) I (like the cat in the) (park) on the (hill).

References

1. S. Maeda, "A Characterization of Fundamental Frequency Contours of Speech," Quarterly Progress Report No. 114, Research Laboratory of Electronics, M. I. T., July 15, 1974, pp. 193-211.
2. H. Hirose, "Electromyography of the Articulatory Muscles: Current Instrumentation and Techniques," Haskins Laboratories Status Report on Speech Research SR-25/26, 1971, pp. 73-86.
3. D. K. Port, "The EMG Data System," Haskins Laboratories Status Report on Speech Research SR-25/26, 1971, pp. 67-72.
4. D. K. Port, "Computer Processing of EMG Signals at Haskins Laboratories," Haskins Laboratories Status Report on Speech Research SR-33, 1971, pp. 173-183.
5. J. E. Atkinson, "Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency," Ph. D. Thesis, University of Connecticut, 1973.
6. R. Collier, "Laryngeal Tension, Subglottal Pressure and the Control of Pitch in Speech" (to be published in Haskins Laboratories Status Report on Speech Research).
7. J. Ohala, "Aspects of the Control and Production of Speech," Working Papers in Phonetics, University of California, Los Angeles, 1970, pp. 1-152.
8. Z. Shimada and H. Hirose, "Physiological Correlates of Japanese Accent Patterns," Annual Bulletin, Research Institute Logopedics Phoniatorics, Tokyo University, Vol. 5, pp. 41-49, 1971.
9. W. A. Lea, "Segmental and Suprasegmental Influences on Fundamental Frequency Contours," Consonant Types and Tone, Southern California Occasional Papers in Linguistics, No. 1, Los Angeles, California, 1973, pp. 17-70.
10. M. Halle and K. N. Stevens, "A Note on Laryngeal Features," Quarterly Progress Report No. 101, Research Laboratory of Electronics, M. I. T., April 15, 1971, pp. 198-213.
11. H. Hirose, "An Electromyographic Study of Laryngeal Adjustments during Speech Articulation: A Preliminary Report," Haskins Laboratories Status Report on Speech Research SR-25/26, 1971, pp. 107-116.
12. H. Hirose and T. Gay, "The Activity of the Intrinsic Laryngeal Muscles in Voicing Control: An Electromyographic Study," Phonetica 25, 140-164 (1972).

C. ACOUSTIC CHARACTERISTICS OF VOWEL NASALIZATION

National Institutes of Health (Grant 2 RO1 NS04332-11)

René Carré

1. Introduction

The effects of vowel nasalization have been described previously[1-3] and the theory of such phenomena has been demonstrated.[4, 5] But acoustical measurements stating precisely the amount of vowel modification when nasalization occurs have not been reported.

In order to obtain such measurements, a set of CVCV utterances read by 5 American male speakers has been analyzed. The consonant C comprised nasal consonants /m, n, ŋ/ and voiced stops /b, d, g/. The vowel V was one of 7 vowels /i, e, ɛ, æ, a, o, u/. Nasalization occurs generally in vowels positioned close to, or more often between, two nasal consonants. Vowels between two voiced-stop consonants are not generally nasalized.

The middle of the first vowel was analyzed by using linear prediction techniques[6] and cepstral prediction techniques[7] (to detect frequencies of spectral zeros). A measure of the degree of nasalization was obtained from the output of a small accelerometer attached to one nostril.

2. Results and Discussion

As an example, formant frequencies and bandwidths for the vowel /ɛ/ are given in Table XVII-1, together with standard deviations. These data were determined from linear prediction techniques, with 13 predictor coefficients used. Also listed are ratios of nasalized vowel formant frequencies (in m, n, ŋ, context) to nonnasalized vowel formant frequencies (in b, d, g, context), and similar ratios for the formant bandwidths. The results show that these coefficients depend on the formant number and the vowel. The standard deviations are usually greater for the nasalized vowels, which suggests that the amount of nasalization is a speaker characteristic.

The bandwidth of the first formant is much greater for the nasalized vowel. Bandwidths are shown for all vowels in Fig. XVII-12. The large standard deviations of ΔF_1 for nasalized vowels can be due to speaker differences and to the technique used to compute these bandwidths. As an example, results obtained for the vowel /i/ read by each of 5 speakers are listed Table XVII-2. Large individual differences can be observed. Table XVII-1 shows that the second-formant bandwidth depends on the vowel but is not really affected when nasalization occurs.

The results obtained with 18 predictor coefficients are sometimes different and appear to be more accurate. The bandwidths are generally smaller, and nasal formants can always be detected in nasalized vowels even when oral and nasal formants are close together. The nasal formant frequencies depend on the vowel and the speaker. These

Table XVII-1. Formant frequencies and bandwidths with standard deviations for the nonnasalized and nasalized vowel /ɛ/. K is the ratio of the two corresponding frequencies.

	F_1 (σ)	ΔF_1 (σ)	F_2 (σ)	ΔF_2 (σ)	F_3 (σ)	ΔF_3 (σ)
ɛ (b, d, g)	537 (49)	54 (19)	1763 (138)	153 (105)	1587 (225)	206 (138)
ɛ (m, n, η)	567 (61)	180 (120)	1817 (139)	144 (40)	2609 (167)	169 (166)
Ratio (K)	1.06	3.33	1.03	0.94	1.01	0.82

Table XVII-2. First formant bandwidth for the vowel /i/ for 5 speakers.

Speaker	ΔF_1 mimi	ΔF_1 nini	ΔF_1 ηiηi
1	156 Hz	183 Hz	147 Hz
2	51	48	54
3	112	180	294
4	41	45	72
5	134	113	275

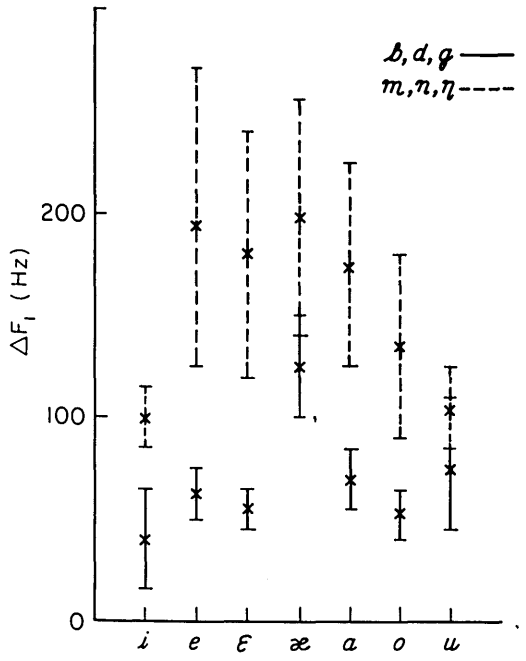


Fig. XVII-12.

First formant bandwidth and its standard deviation for 7 nonnasalized vowels (in b, d, g context) and nasalized vowels (in m, n, η context).

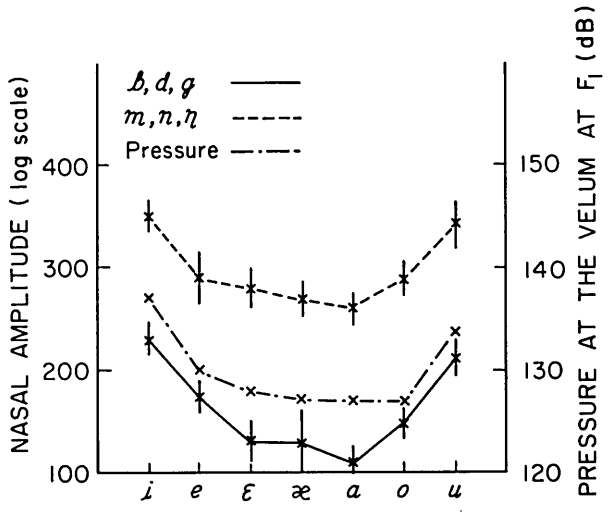


Fig. XVII-13. Left: nasal amplitude for nasalized and nonnasalized vowels. Right: pressure level at the velum at the first formant frequency. (One unit on the amplitude scale is approximately 1/10 dB.)

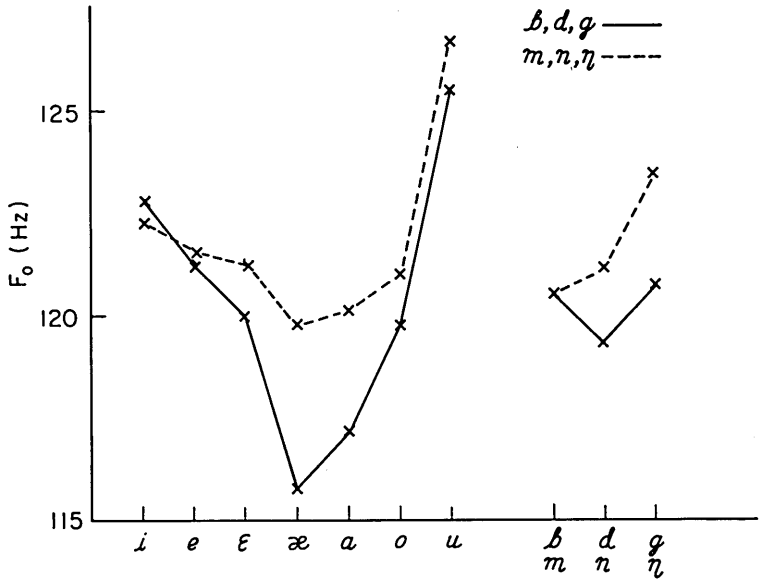


Fig. XVII-14. Fundamental frequency for nasalized and nonnasalized vowels.

frequencies are around 1000 Hz and 2000 Hz, but these mean values vary so much with the speaker that they are not very significant. The zero frequencies detected by using cepstral prediction techniques are around 1300 Hz and also depend on the vowel and the speaker.

The amplitudes of the nasal accelerometer output for nasal and nonnasal vowels are shown in Fig. XVII-13. The amplitude is greater for the nasalized vowels /i/ and /u/. This result seems to be due to the higher pressure levels near the velum at the first formant frequency. These pressure levels are shown here from measurements on an electrical analog for the vocal tract. For the nonnasalized vowels, presumably, there is transmission of sound energy from the oral cavity through the velum, the transmission being more efficient for vowels with a low first formant.

The fundamental frequency is generally higher with nasalized vowels than with nonnasal vowels (Fig. XVII-14). Results for nonnasalized vowels fit well with the data of House and Fairbanks.[8] But no definitive explanation can be advanced for the big gap between the nasalized and nonnasalized vowels /æ/ and /a/. Among other features, the durations of the nasalized vowels are slightly longer than those of the nonnasal vowels, and their intensities are almost the same.

3. Conclusion

It seems that these results may be used in speech recognition, with the nasalization feature detected by means of bandwidth measurements. But the amount of nasalization depends on the speaker. Thus when nasalization occurs special care must be taken not to detect the nasal formant in place of the second formant. Then before matching special corrections can be made according to the class of the vowel and the speaker.

References

1. A. S. House and K. N. Stevens, "Analog Studies of the Nasalization of Vowels," J. Speech and Hearing Disord. 21, 218-232 (1956).

2. A. S. House, "Analog Studies of Nasal Consonants," J. Speech and Hearing Disord. 22, 190-204 (1957).

3. C. G. M. Fant, Acoustical Theory of Speech Production (Mouton and Co., 's-Gravenhage, The Netherlands, 1960).

4. O. Fujimura, "Spectra of Nasalized Vowels," Quarterly Progress Report No. 58, Research Laboratory of Electronics, M. I. T., July 15, 1960, pp. 214-218.

5. O. Fujimura, "Analysis of Nasal Consonants," J. Acoust. Soc. Am. 34, 1865-1875 (1962).

6. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am. 50, 637-665 (1971).

7. A. V. Oppenheim and J. M. Tribolet, "Pole-Zero Modeling Using Cepstral Prediction," Quarterly Progress Report No. 111, Research Laboratory of Electronics, M. I. T., January 15, 1974, pp. 157-159.

8. A. S. House and G. Fairbanks, "The Influence of Consonantal Environment upon the Secondary Acoustic Characteristics of Vowels," J. Acoust. Soc. Am. 25, 105-113 (1953).