Synthesizing intonation

Janet Pierrehumbert

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 4 March 1980; accepted for publication 8 June 1981)

This paper describes a computer program for synthesizing fundamental frequency (*F*0) contours for English, and sketches the theory which underlies it. The *F*0 contour is described as a series of targets within an envelope specifying *F*0 range; the *F*0 contour between targets is computed by transition rules. The use of nonmonotonic transitions permits a sparser specification of the contour than has been possible in most previous frameworks. The program generates a good synthesis of neutral declarative intonation. Unlike most previous *F*0 synthesis programs, it can also be used to synthesize a variety of non-neutral intonation patterns.

PACS numbers: 43.70.Jt, 43.70.Qa

INTRODUCTION

Intonation, which is a major determinant of the patterns of fundamental frequency (*F*0) seen in speech, has attracted attention from a number of points of view. Linguists have been interested in characterizing what different intonation patterns exist, what sorts of meaning are conveyed by these patterns, and how the temporal relation of the *F*0 contour to the speech segments is governed by stress and syntax. Psycholinguists have begun to examine intonation as an indirect source of evidence about what the units of planning in speech are and how parsing is accomplished. From a more applied point of view, it has been suggested that intelligent use of information carried by *F*0 could improve automatic speech recognition. In particular, Lea (1972) developed a phrase-boundary detector based on *F*0, and Vives *et al.* (1977) describe a component of a speech recognition system for French which uses *F*0 to screen word candidates put forward by a segmental analyzer. Lastly, it is clear that intonation plays an important part in the intelligibility and naturalness of synthetic speech (Olive and Nakatani, 1974; and Nooteboom *et al.*, 1976). This means that an adequate intonation synthesis program is an important prerequisite to any speech synthesizer with practical applicability for extended utterances.

The work described here was primarily concerned with synthesis of intonation. The aim in doing this work was to use analysis-by-synthesis to make progress towards a full model of English intonation. By contrast, the main tradition of work on intonation synthesis, exemplified by Vaissière (1971), Olive (1974), and O'Shaughnessy (1976), has had a more practical bent, having been concerned with developing a program to provide *F*0 contours for a text-to-speech system. While the work reported here resulted in a proposal for handling this problem, the difference in major emphasis between the present work and previous work has a number of ramifications which should be noted.

The input to the *F*0 component of a text-to-speech system is typically underspecified in the sense that it does not provide information about all of the linguistic factors which are known to play a part in governing the *F*0 contour. A reading machine for the blind, for example, must operate on standard English orthography, which provides no information about the location of stressed syllables and only marginal information about parsing. Olive's *F*0 synthesis program (Olive, 1974) was designed for a word concatenation synthesizer in which the location of lexical stress was unavailable; and the MITalk system described in Allen *et al.* (1979) provides some parsing information to O'Shaughnessy's *F*0 program, but this information is in some cases incomplete or incorrect. More broadly, in any text-to-speech system, the computer must assign an *F*0 contour without understanding what the text is saying. Enormous advances in artificial intelligence will be needed before we can expect a computer to mimic people in using a variety of intonation patterns to express a variety of relations between the current utterance and the sense of the previous discourse. For example, Liberman and Sag (1974) describe an intonation pattern which may be used when the speaker is contradicting his interlocutor; at present, however, it is impossible to program a reading machine which would use this contour appropriately in artistic renditions of plays. Because text-to-speech systems cannot make appropriate choices among the intonation patterns of English, they are programmed to generate only one kind of intonation per sentence—what might be called neutral declarative intonation. As a result, speech communication research on intonation has concentrated on the problem of approximating this particular kind of intonation; there has been relatively little attention to other patterns which are common in natural discourse.

The underspecification of input to the *F*0 program for a text-to-speech system has an additional consequence for the form of the output: Insofar as possible, the system should mimic natural speech, but because this is not always possible, the system should also minimize the abrasiveness of deviations. The second requirement may in principle override the first, resulting in a system which deviates systematically from a true model, even in its limited sphere. For example, Olive's program, which does not have access to the location of stressed syllables, cannot locate *F*0 peaks on the stressed syllables, as would be normal in neutral declarative intonation. Instead, the program places on each lexical word a broader *F*0 peak than would be found in normal speech. The inevitable discrepancies between peak location and stress location are thus less noticeable than they would be if the program copied faithfully the peak shapes found in natural speech.

 0001-4966/81/100985-11$00.80

The researcher whose aim is to model intonation rather than to provide *F0* for a text-to-speech system is not constrained by underspecified input. Rather, the problem is to design an input which encodes appropriately the knowledge about a sentence that a speaker would use in computing its intonation; the synthesis program itself embodies a set of rules for translating this knowledge into an *F0* contour. This problem may be compared to the problem of designing the input for the segmental component of a synthesis-by-rule system. Ultimately, we might wish the computer program to decide what to say, construct a sentence, look up the words in its lexicon and retrieve their transcriptions, and pass these transcriptions to a routine which converts them into sound. In the short run, however, we work on such a complete model one module at a time. Typically, the researcher selects the words and passes a phonological or orthographic transcription to a synthesis routine. Similarly, the program described here accepts a transcription of an intonation pattern and converts it into an *F0* contour for an utterance. The utterance which carries the *F0* contour may be synthesized from linear predictor data,[1] or by rule from a segmental transcription.

In designing a transcription system for an intonation model, we would wish to keep two goals in mind. First, the transcription system must permit us to duplicate the linguistic distinctions that a speaker can make. A segmental synthesizer which uses Arpabet transcription allows us to duplicate the difference between the pronoun "mine" and the author "Mann" by typing AY versus AA; similarly, the system adopted for transcribing intonation should make it possible to duplicate the intonation of a given sentence said in various ways, e.g., to convey information; to ask a yes/no question; or to express shocked surprise. Secondly, we want to leave out of the transcription things that we are able to predict by rule. A segmental synthesizer which incorporates rules for unstressed vowel reduction, dental flapping, and /l/-velarization represents more understanding than one which requires a detailed phonetic transcription that provides such information; a synthesizer which requires a phonetic transcription in turn represents progress over one which takes as input time functions for excitation source characteristics and vocal tract resonances.

The intonation synthesis program which will be described below takes a phonetic transcription of the intonation pattern as input. This means that the program incorporates a number of conclusions about what the units of intonational description are and how these units are realized as continuous *F0* contours. Sections I and II sketch these conclusions and how they are incorporated into the program; Sec. IV reviews how they are supported by the literature on speech production and perception. At the same time, the program still takes as input some information which a more complete model would handle by rule. Section III proposes a more complete description of neutral declarative intonation, and describes the routine which is used in the text-to-speech system described in Olive and Liberman (1979) to supply input to the program described here. When the program is not driven by this routine, but rather by input from a file, it can compute additional intonation patterns and serves as a tool for investigating more complete models of intonation.

I. BACKGROUND ASSUMPTIONS

A number of conclusions about intonation which are relevant to the *F0* synthesis program may be illustrated by considering Fig. 1. This figure shows a natural *F0* contour for the sentence "In November, the region's weather was unusually dry." Following Lehiste and Peterson (1961), Öhman (1967), Vaissière (1971), Lea (1972), and others, we distinguish two contributions to this *F0* contour. The prosody, or choice of stress, phrasing, and intonation pattern, is responsible for the overall shape of the contour. In addition, the speech segments introduce local *F0* perturbations. In particular, high vowels tend to raise *F0*; unvoiced obstruents raise the pitch at the onset of a following vowel; and voiced obstruents and glottal stops are associated with a dip in *F0*. These effects are responsible for the fine variations in *F0* in Figs. 1–4. Data on the magnitude of the effects in a number of contexts may be found in Lea (1972). Bruce (1977) also gives extensive references. However, the interaction of these effects with prosody in running speech has not been sufficiently researched to permit construction of a quantitative model. Thus, the following discussion will be concerned only with how to model prosodic effects on *F0*. While *F0* contours lacking segmental effects are noticeably smoother than those of natural speech, this smoothness appears to impair only marginally one's ability to judge the correctness of the prosody. Thus the lack of segmental effects on *F0* in the model is not a major handicap to the present enterprise.

Having set aside the segmental influences on *F0*, we would like to distinguish two aspects of the prosodic contribution to *F0*: the pitch pattern and the pitch range. Writers since Trager and Smith (1951) have recognized that it is possible to produce the same in-

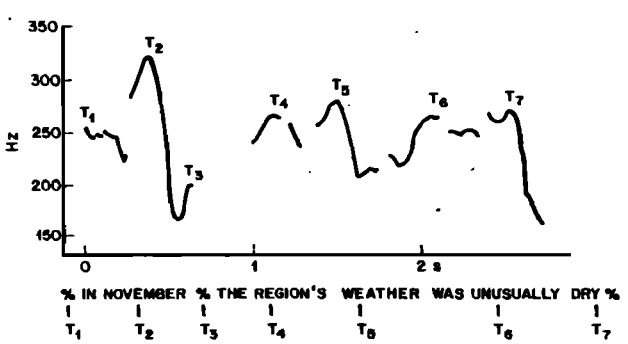

FIG. 1. *F0* contour for an utterance of the sentence "In November, the region's weather was unusually dry." *T1* through *T7* are points in the contour which are interpreted as *F0* targets in the present synthesis rules. The labeling under the contour indicates how these targets are aligned with syllables and phrase boundaries (%).

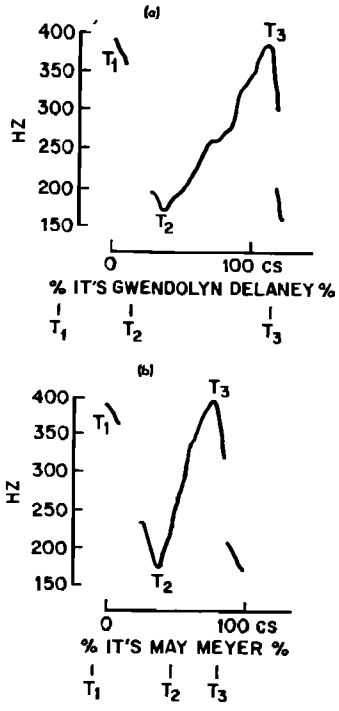

FIG. 2. *F*0 contours for the sentences "It's Gwendolyn Delaney!" and "It's May Meyer!," produced using the surprise/redundancy intonation pattern described in Sag and Liberman (1975). As in Fig. 1, subscripted *T*'s are used to indicate what points in the contour are viewed as targets and how these points are aligned with the text.

tonation pattern in different pitch ranges. The reader can persuade himself that this is true by calling out the name of someone he imagines to be close by or far away. A somewhat more subtle point is that the pitch range varies systematically within the utterance; in many studies of neutral declarative intonation, it has been found that the pitch range narrows and drifts downwards over the course of a major phrase [for example, see Maeda (1976), O'Shaughnessy (1976), Cooper and Sorensen (1977), and Sorensen and Cooper (1980)]. This tendency, which is referred to as the declination effect, can be seen in the overall trend of the *F*0 contour in Fig. 1. Two possible interpretations of this trend suggest themselves. Either the stressed syllables which have peaks become less and less prominent over the course of the sentence, or else a peak does not have to be as high later in the sentence as it was earlier in order to express a given degree of prominence. Perception experiments reported in Pierrehumbert (1979b) bear out the second interpretation. For two peaks in a sentence to sound equally high, the first in general had to have a higher *F*0. When the two peaks had the same *F*0, the contour sounded well-formed, but the second was perceived as higher. These results suggest that declination is involved at an implicit level in computing what *F*0 levels correspond to what degree of prominence, even when the peaks later in the utterance are sufficiently prominent that the *F*0 contour does not exhibit on the surface the overall declining shape of the contour in Fig. 1. This account receives further support from an experiment reported in Liberman and Pierrehumbert (1979) and Pierrehumbert (1980). This experiment investigated how the *F*0 value of a more prominent peak was related to that of a less prominent one under changes in pitch range and order. In fitting a quantitative model to these data, it was found that the two peaks were in a constant ratio if they were scaled by an implicit declining baseline which was fixed for each speaker. The declination was implicit in the sense that it was not evident in any single *F*0 contour; declination in the upper part of the range was obscured by the co-occurrent prominence differences, while declination of the baseline was not evident because the patterns selected for study touched the baseline only once. However, the baseline was still well defined: For each speaker, separate estimates of the baseline based on peak relations and on low points in the aggregate of *F*0 contours came out to the same value.

In view of these results, the *F*0 synthesis program permits the researcher to assign a pitch range func-

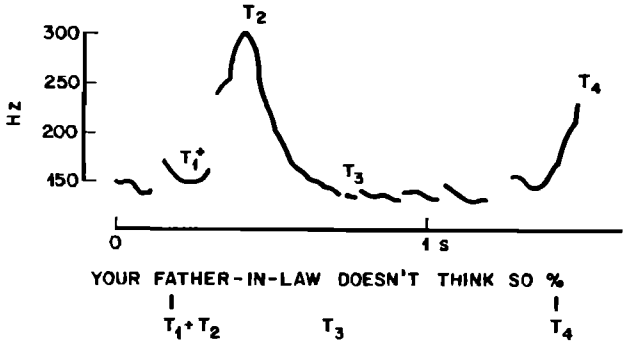

FIG. 3. An *F*0 contour for "Your father-in-law doesn't think so," produced with the nuclear stress on "father." A two-target pitch accent on "father" is followed by two additional targets *T*3 and *T*4, which mark the phrase as a nonterminal declarative. In Sec. II, we will see that *T*3 would be supplied by the program in a synthesis of this intonation pattern, and thus would not be marked in the input.

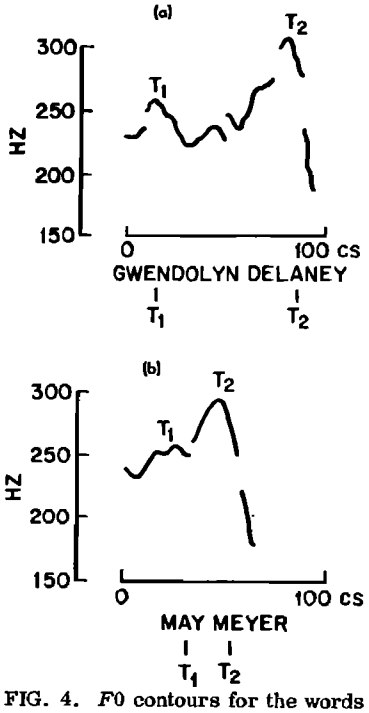

FIG. 4. *F*0 contours for the words "Gwendolyn Delaney" and "May Meyer" excised from the context "I think you should discuss it with _____." Subscripted *T*'s indicate two points which are to be viewed as *F*0 targets. When the two targets are far apart, as in (a), the *F*0 contour sags in between; when they are sufficiently close together, as in (b), the sagging disappears.

tion to each phrase. Values at a particular point, as indicated by the transcription, are then situated within the current pitch range. The results of the experiment just cited support a model in which declination throughout the pitch range is determined by the baseline declination. This restrictive hypothesis has not been incorporated into the *F0* synthesis program, pending further confirmation. Instead, the program permits a topline and a baseline to be assigned independently. The topline and the baseline are the envelope for the *F0* contour; the contour itself is described in terms of fractions of the distance from the baseline to the topline.

As is clear in Fig. 1, *F0* varies continuously in natural speech. Just as it is useful to view the continuously varying vocal tract resonances as the implementation of a string of discrete speech segments, it is also useful to view the continuously varying *F0* contour as implementing a series of discrete elements. There have been a number of proposals about what these discrete elements are. Öhman (1967) analyzes Swedish *F0* contours in terms of impulses fed to a linear filter. Each word receives one impulse, and all impulses are in the same direction. A related model for *F0* in Japanese is presented in Fujisaki and Nagashima (1969). Bolinger (1951), 't Hart and Cohen (1973), O'Shaughnessy (1976), and Clark (1978) propose theories under which the contour in Fig. 1 would be analyzed as a series of instructions to raise or lower the *F0*. A third school of thought, represented by Pike (1945), Trager and Smith (1951), Liberman (1975), and Bruce (1977), would analyze the contour as a series of target values which are connected together by transitional functions. The work presented here takes this third approach. For example, both of the *F0* contours shown in Fig. 2 have a high target value at the onset (*T1*), a low target on the first stressed syllable (*T2*), and a high target on the second stressed syllable (*T3*). The slope of the contour between *T2* and *T3* is greater in Fig. 2(b) than in Fig. 2(a) because there is less time to connect the first to the second in Fig. 2(b) than in Fig. 2(a).

It is clear that any particular *F0* contour could be generated under any of these three approaches. Thus, arguments for one approach over another must be made on the basis of regularities in the entire system of intonation. Such a case for analyzing English intonation in terms of target values is made in Pierrehumbert (1980). There, the phrasal tunes of English are decomposed into targets associated with stressed syllables and targets associated with the margins of the phrase. The tonal marking of a stressed syllable may be either a single target (like the tonal marking of the stressed syllables in Fig. 2), or a sequence of two targets. Such a tonal marking is called a pitch accent. A stressed syllable may lack a pitch accent, in which case its *F0* contour is determined by the transition between adjacent targets just as if it were unstressed. The targets associated with the margins of the phrase are the boundary tones which control the *F0* at the onset and the offset of the phrase, and an additional target which falls in between the pitch accent on the nuclear, or main, stress of the phrase, and the phrase-final boundary tone. In Fig. 2, the phrase-initial boundary tone lends a note of vivacity to the utterance; a low onset would be more neutral. In Fig. 3, a two target (low-high) pitch accent on "father" is followed by a low target and then a high boundary tone. This low-high phrase final configuration is often used to indicate that the speaker intends to continue; it is also used to imply that the listener may have something to add.

In Pierrehumbert (1980), a case for this approach is made on the basis of observations about what intonation patterns occur in English and what do not, what *F0* configurations count as instances of the same pattern, and which syllables receive pitch accents. A comparable case for analyzing Swedish *F0* contours in terms of target values is made in Bruce (1977).

An important feature of the model proposed here is that the transition between two targets is not always monotonic. When one target is at the bottom of the pitch range, as in Fig. 2, the transition is monotonic. However, when neither target is near the baseline, a sagging transition is used. For example, to synthesize the *F0* contour covering the words "the region's weather" in Fig. 1, targets *T4* and *T5* are assigned to the stressed syllables, but the valley in between is generated by the transition rules. The resulting synthetic contour can be seen in Fig. 5. Under previous target theories of intonation, an additional target corresponding to the bottom of the valley would have to be supplied in the input. The amount of sagging between two targets depends on their separation in time and can, for sufficiently close targets, be zero. We take the *F0* contour on "May Meyer" in Fig. 4(b), for example, to have a target on "May" and one on "Meyer"; these targets are close together and the transition is virtually monotonic. Note that under these assumptions, "May Meyer" in Fig. 4(b) has the same intonation as "Gwendolyn Delaney" in Fig. 4(a): Both contours have a target in the middle of the pitch range associated with the first stressed syllable, and a target at the top of the pitch range associated with the second stressed syllable. The claim that these are instances of the same intonation pattern is supported by a strong gen-

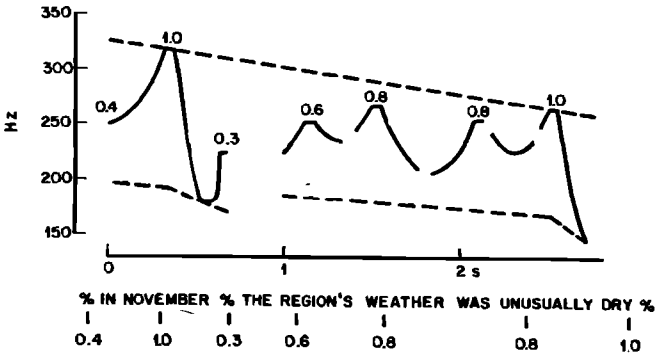

FIG. 5. Synthetic version of the *F0* contour in Fig. 1, generated using formula (1) as input to the synthesis program. The dashed lines indicate the current *F0* range at each point in time. *F0* targets and their alignment with the text are indicated by decimal numbers, which are interpreted as a proportion of the way from the bottom to the top of the *F0* range.

eralization in O'Shaughnessy (1976). The *F*0 contours that he gives include 89 instances in which two targets are related as in Fig. 4. Of the 51 which have a dip, as in Fig. 4(a), 48 have one or more unstressed syllables between the two targets. Of the 38 which have no dip, as in Fig. 4(b), 36 have no unstressed syllables between the two targets. Thus, one of the major motivations for the proposed transition rules is that they define equivalence classes among *F*0 contours in this way.

II. HOW THE PROGRAM WORKS

The preceding section suggested that the *F*0 contour may be described as a series of target values which are connected together by transition rules. The target values are expressed as locations within the current pitch range, which varies as a function of time. Which syllables within the phrase are assigned a target depends on the stress pattern; targets may also be assigned to phase boundaries. This section discusses how the transcription system and implementation rules used in the intonation synthesis program incorporate these assumptions.

The input to the program is a string of phonemes, annotated with durations, phrase boundaries, and target levels. Equation (1) shows the input to generate a synthetic version, Fig. 5, of the natural contour shown in Fig. 1.

```
{325 260 195 165}(0.4) SIL 8 ih 4 n 3* n 4 ow 7
v 9 eh 10(1.0) m 8 b 4 er 13%(0.3) SIL 34
*dh 2 ax 2* r 5 iy 7(0.6) jh 10 en 8 z 10*
w 3 eh 7(0.8) dh 4 er 8* w 6 ax 5 z 7*                    (1)
ax 4 n 9 yu 16(0.8) zh 6 ax 5 l 4 iy 9*
d 8 r 5 ai 24(1.0)%
```

The segmental transcription is an Arpabet. The numbers enclosed in curly braces mark the major phrase boundary and generate the pitch range function shown in the figure. % is the minor phrase boundary. The intonation program does not make any use of word boundaries, but word boundaries may be marked, as here, if the input will also be used by the segmental synthesis rules.[2] Target values, which are surrounded by parentheses, are allowed to take values from 0–1. 0 is on the baseline; 1 is at the current top of the pitch range; intermediate values are placed in between on a linear scale of the fundamental frequency (in Hz). Each target value generates a 6-cs-level section in the *F*0 contour. The lowest value the *F*0 reaches between two targets is determined by rule, and may correspond to the lower of the targets; a quadratic function is then fit on the basis of the targets and this minimum value. Quadratics have been found to be a satisfactory choice, but undoubtedly they are not the only possible satisfactory choice since the ear seems to be relatively insensitive to the shape of the curve between targets. This is not surprising, since it is the target points which carry the important linguistic information, and not the contour in between.

Details of the scheme of implementation depend on a more theoretical analysis of the targets. Work in linguistics has suggested that the target values, which here go from 0–1, are divided into linguistic categories, much as the possible continuum from high-to-low front vowels is divided into the English front vowels. The traditional number of different categories for target level in English is four (Trager and Smith, 1951; Pike, 1945; Liberman, 1975). However, Pierrehumbert (1980) reduced the number to two, high and low (hereafter *H* and *L*). Allophonic rules were shown to account for the *F*0 configurations which had seemed to require additional tonal categories. Reducing the number of tones not only simplifies the description, but also makes it possible to answer objections raised in Bolinger (1951) and Ladd (1978) to analyzing intonation patterns as sequences of target values. There are three major phonetic differences between *L* and *H*. First, *L* is lower in the speaker's range than *H* would be at the same location. The program described here takes any target value less than 0.2 to represent *L*, although it is clear that this is a simplification. Second, if the speaker increases the emphasis on a *H* pitch accent, the target *F*0 value is higher, while increasing the emphasis on a *L* pitch accent lowers the target value. The effect of relative prominence on the relative *F*0 values of *H* targets can account for the target relations in Fig. 4. *T*2 is higher than *T*1 because it has the main phrase stress, whereas *T*1 has a subordinated stress. Lastly, the transition between two *H*'s displays the sagging illustrated in Fig. 4. The transition between *L* and another tone is monotonic, as in Fig. 2. For this reason the implementation rules are sensitive to whether a target is *H* or *L*. In general, they realize *H*'s as peaks and *L*'s as nonpeaks.

Between two *H* targets, the sagging principle illustrated in Fig. 4 is applied. The properties of the rules for computing the minimum *F*0 between two targets were worked out on the basis of the successes and weaknesses of a previous synthesis program described in Pierrehumbert (1979a). As in the old model, targets which are well separated in time are implemented as separate peaks, with the slope of rise and fall surrounding the peak increasing with target height. This principle is illustrated in Fig. 6. When the peaks were closer together, the old model had the *F*0 track the fall as determined by the height of the first peak until it intersected the rise as determined by the height of the second peak. This formulation frequently resulted in excessive *F*0 movement. In these circumstances, it would appear, speakers adjust the *F*0 movement immediately out of the first target according to where the next target is. Figure 7 illustrates how the implementation system used here incorporates this effect for a moderate separation of peaks. As two targets are placed closer and closer together, this effect dominates more and more over the independent effect of each target's height on the rate of *F*0 change surrounding the target. In particular, the *F*0 dip between two targets of the same height which are very close together (under 20 cs apart) is small, and relatively unaffected by the height of the targets.

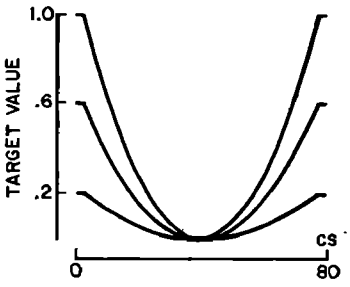

FIG. 6. Superimposed *F*0 contours showing transitions computed between two targets which are 80 cs apart. Any path through the graph is an output of the system. In every case, the *F*0 contour reaches the baseline 40 cs after the first target.

The computational form which was adopted so that the *F*0 transitions would have these properties is as follows: When the subroutine computing the *F*0 minimum between two targets is entered, the targets have already been translated into *F*0 values on the basis of the value of the baseline and topline at the time of each target. When these two target *F*0 values are the same, the *F*0 minimum is found as a fraction of the distance between the lower of two baseline values at the targets and the *F*0 value of the targets. This fraction is a function of the separation in time of the targets. The function is two-piece linear: the first piece (running from 0–20 cs) is specified by Eq. (2) and the second piece (running from 20–80 cs) is specified by Eq. (3).

$$F = 1 - (T^* \ 0.005). \tag{2}$$

$$F = 0.9 - [(T - 20)^* \ 0.015]. \tag{3}$$

In both (2) and (3), F is the computed fraction and T is the separation in time of the two targets in centiseconds. When T is greater than 80, the two targets are implemented separately; the fall from the first takes up 40 cs, and then the *F*0 tracks the baseline until it begins to rise 40 cs before the second. When the two targets are not equal, an F is computed by Eq. (1) on the basis of their separation in time. F' is then computed as F scaled by $(TL-B)/(TH-B)$, where TH is the higher of the two targets, TL is the lower of the two, and B is the lower of the two baseline values at the targets. The final F is the geometric mean of F and F'. If the computed F comes out higher than the lower target, it is reset to the lower target.

The locations of the targets and the value of F analytically determine the parabola which is fit between the two targets.

When one or both of the targets is L, the rules for computing a sagging transition which were just described are set aside; an *F*0 minimum for the parabolic transition is set at the time and height of the lower target. Thus, in a L–H sequence, a rise starts immediately after the L regardless of how far it is from the H. In a L–L sequence, one L may of course correspond to a higher *F*0 than the other, but the *F*0 contour takes a monotonic course in between, never dipping below the lower value.

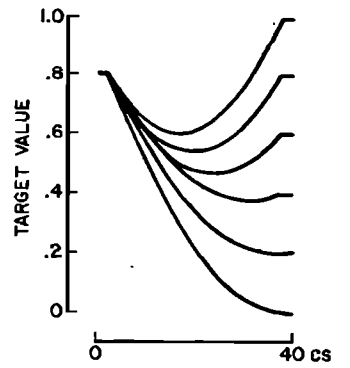

FIG. 7. Superimposed *F*0 contours showing transitions computed between two targets which are 40 cs apart.

A second theoretical point which plays a part in the synthesis program is the distinction between the nuclear pitch accent, which falls on the main phrase stress, and the prenuclear pitch accents. As many authors have observed, the nuclear accent is the last pitch accent in the phrase (O'Connor and Arnold, 1961; Crystal, 1969; Vanderslice and Ladefoged, 1972; Ashby, 1978). Although there may be stressed syllables after the nuclear stress, they cannot have pitch accents. In the framework presented here, their *F*0 contours are generated from the two extra targets marking the end of the phrase which were introduced above. Where the nuclear stress is on the last syllable, these two extra targets are also crowded onto the same syllable.

The interaction of the extra targets with the nuclear pitch accent generates effects which must be replicated for the synthesis to sound natural. These effects have been most thoroughly studied in neutral declarative intonation, where L follows a nuclear H accent. It has been found that the H often occurs earlier in the stressed syllable than a prenuclear H would. The fall from H to L must be steeper than falls elsewhere for the pitch accent to sound nuclear (Olive, 1974). When there is a L phrase-final boundary tone, it is produced lower than an extrapolation of low points earlier in the phrase would predict (Maeda, 1976; and Mattingly, 1968).

In order to generate these phonetic details, the program treats the nuclear accent as a special case. Since the nuclear accent is the last accent in each phrase, it can be identified without elaboration of the input. The location of the nuclear H was handled in Pierrehumbert (1979a) in accordance with the observation in Ashby (1978) that the nuclear H occurs a fixed distance into the vowel. This meant that the peak was a relatively small percentage of the way into a phrase-final stressed syllable, which is long; and a greater percentage of the way into a stressed syllable which is not phrase-final, and is therefore shorter. Further experimentation with the synthesis program showed that it was impossible to find a fixed distance controlling peak placement which was suitable for all words. To leave room for the nuclear fall on a word like "bit," the peak had to be placed a very short distance into the vowel; and when this short distance was used to place the peak in a disyllabic word with a long stressed first syllable, such as "rival," the *F*0 was then much too low by the onset of the second syllable. As a result, the present program places the peak early in the syllable only in the case of phrase-final nuclear syllables. In order to guarantee a steep nuclear fall, the program still incorporates Ashby's observation that the fall occurs in a fixed amount of time, here 20 cs; however, this rule has resulted in some anomalies and further

work on where the *L* is located after *H* is in progress. The program also makes the nuclear fall go below the baseline for the phrase as a whole by lowering a trapdoor in the baseline after the nuclear stress, as can be seen in Fig. 5. What the program does, then, is to generate for each nuclear *H* a following *L* which has a particular placement in time and pitch. We interpret the fall theoretically as a sequence of *H* and *L*, but only the *H* has to be specified in the input.

A target value for a *H* boundary tone may be specified optionally. If it is missing, the program generates an *L* boundary tone automatically.

III. NEUTRAL DECLARATIVE INTONATION

The *F*0 contour in Fig. 1 exemplified neutral declarative intonation. We propose the following theoretical description for this type of intonation. The pitch accents are all *H*. When the phrase is terminal, the phrase-final tones are typically *L–L*. If it is nonterminal, they are typically *L–H*. The target values to be input to the program may be determined from this description by using the phrasal stress subordination. As noted above, the *H* on a syllable with higher stress corresponds to a higher target value within the current pitch range.[3]

This general idea can be made more precise using the formalism for representing stress developed in Liberman (1975) and Liberman and Prince (1977). We discuss a single example here, and refer the reader to the original sources for a more general account. To compute the stress contour for a phrase, we begin by constructing a syntactic tree for it. Figure 8 shows the tree for the second phrase in Fig. 1. Function words which are unstressed are assumed here to be cliticized, and so they are not entered as words in the tree. Syntactic labeling (e.g., noun phrase, verb phrase) is omitted to make room for node labeling referring to stress relationships. In general, stronger stress is assigned to the right of two sister nodes in a syntactic phrase by Chomsky and Halle's Nuclear Stress Rule (Chomsky and Halle, 1968). This relationship is indicated by labeling the right node "*s*" (strong) and the left node "*w*" (weak). Here, "dry" receives stronger stress than "unusually"; and "was unusually dry," taken as a constituent, receives stronger stress than "the region's weather." We recognize that focus may override the assignment of stress which these rules would predict. For instance, if someone has already remarked that the weather was dry, one might say "Yes, it is *unusually* dry." In this case, "unusually" would of course have stronger stress than "dry," and it would be labeled with "*s*" while "dry" would be labeled with "*w*."

At this point, it will be useful to assign a measure of absolute prominence to each stressed syllable in a way which reflects the stress relationships in the tree in Fig. 8. A numbering scheme which accomplishes this (in fact, the flattest numbering scheme which gives each "*w*" lower prominence than its sister "*s*") is to count up the total number of *w* nodes dominating the syllable. Adding one to this number yields a standard stress transcription in which 1 is the highest stress and higher numbers designate lower stress. The outcome in this case is 3 2 2 1. It should be noted that this transcription differs from what Chomsky and Halle (1968) would generate for the sample phrase; using their rules, the 2-stress on "unusually" would be downgraded to 3 when the phrase "was unusually dry" is embedded as in Fig. 8. The Liberman and Prince transcription is adopted for the present purposes because informal experimentation suggested that *F*0 peaks should not be downgraded in this fashion. In particular, for an extended right-branching construction such as the one in Fig. 9, the Chomsky–Halle rules generate a stress contour which descends steadily from the first stress through the next-to-last stress: 2 3 4 5 1. The stress contour as determined by adding 1 to the number of *w* nodes dominating each stress would be 2 2 2 2 1; however, a further principle proposed in Liberman (1975) requiring alternation in stress contours predicts that the contour would instead be any of a number of alternating configurations, such as 2 3 2 3 1. In fact, a neutral declarative intonation for this sentence does not have a series of peaks which descend as the Chomsky–Halle stress transcription does, but rather has a series of peaks which alternate in height, with the exact configuration of alternation at the speaker's discretion.

A good synthesis of neutral declarative intonation may be achieved by assigning to a 1-stress a target on the topline; to a 2-stress, a target a third of the way

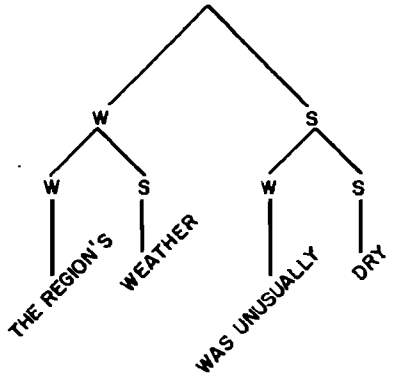

FIG. 8. Metrical tree for the second phrase of the example sentence used in Figs. 1 and 5.

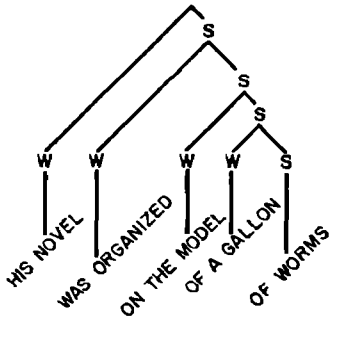

FIG. 9. Metrical tree for an extended right-branching structure.

down from the topline to the baseline (on a linear scale); and to a 3-stress, a target two-thirds of the way down from the topline to the baseline. This principle of target assignment is used in Pierrehumbert (1979a). However, it is clear that human production of intonation displays more latitude than this; we would guess that any perceptually salient difference between two peaks may be used to mark stress subordination in neutral declarative intonation. Subject to the constraint that a stronger stress gets a higher peak than a weaker stress, peak height could then be varied at will to express different degrees of emphasis on any given word. For example, in the sentence in Fig. 1 it would be possible to produce the word "unusually" with a higher or lower peak, according to how newsworthy the speaker feels it is that the dryness is unusual.

A program which generates input for the intonation synthesizer described here was written by Mark Liberman for use with the text-to-speech systems described in Olive and Liberman (1979) and Browman (1979). This program does not have access to a complete parse of a sentence, and so relative prominence cannot be computed by the method just described. Instead, the nuclear stress is assigned to the last content word in each phrase, and receives a target of 1. Prenuclear main word stresses are assigned alternating values of 0.4 and 0.7. The intonation which results from this simple algorithm is quite acceptable, although not surprisingly it sometimes assigns words an unnatural importance in the message.

At this point, we would like to compare the present proposal for synthesizing *F0* contours to other approaches in the literature. 't Hart and Cohen (1973) describe a system for synthesizing Dutch intonation in which a fall is assigned to the last stressed syllable in the phrase, and a rise is assigned to all previous stressed syllables. The *F0* falls off after a rise in order to prepare for the next rise; these nonprominence lending falls are observed to be smaller in real speech than the terminal fall. 't Hart and Cohen's treatment of stressed syllables is related to that proposed for English in Mattingly (1966) and (1968). Mattingly's program generates an increase in *F0* at each prominent (or accented) syllable. A more gradual fall is computed on the syllables which follow up to the next rise. Proposals by Lea (1973) and Klatt (1979) assign to each phrase a hat contour which extends from the first stressed syllable to the last stressed syllable. Peaks for stressed syllables are then added on to this hat contour.

The *F0* contours generated by the present system for neutral declarative intonation have a number of similarities to the contours generated by these systems. Because a prenuclear *H* is located late in the syllable, prenuclear stressed syllables have rising *F0* contours. The nuclear stressed syllable has a fall when the *H* is placed early in the syllable, as it would in Lea's, Klatt's, and 't Hart and Cohen's syntheses. We found that locating the peak early in the nuclear syllable was only successful when the syllable was phrase-final. Mattingly (1966) also treats phrase-final nuclear syllables as a special case. Like the other systems, our system ensures that the terminal fall falls further than earlier falls. Lastly, as Fig. 6 and 7 showed, the rules for computing the transition between two *H* accents take the *F0* down to the baseline only when the targets are exceptionally far apart. Thus, in a typical case, the *F0* contour computed for a series of *H* accents would not fall to the baseline and could be reasonably well approximated by adding peaks onto a phrasal hat contour, as Klatt and Lea propose.

Next to these similarities, there are some important differences. All systems implement *H* pitch accents as peaks by computing a fall and rise; however, ours is the only one in which the upcoming peak affects the *F0* contour from the moment it leaves the last peak. We believe that this lookahead is responsible for a significant improvement when accents are close together. The rules for synthesizing neutral declarative intonation proposed here produce different peak heights on accented syllables, according to their relative prominence. The systems proposed by 't Hart and Cohen, Lea, and Klatt, by contrast, do not make use of stress subordination above the level of the word. Thus, all peaks in effect correspond to targets on the topline in the present system. Experimentation with the present system suggested that this pattern of target assignment produces an adequate *F0* contour for short utterances, but results in a sing-songy effect for texts of any length; subordination among peaks is an important factor in creating an impression that the computer knows what it is saying. Mattingly's program does not refer to phrasal stress subordination, but it may still generate enough alternation in peak height to avoid monotony. Since the amount of rise for a prominence peak is fixed, while the amount of fall following depends on the length of the following material, it appears that peaks would be variously located on the speaker's range.

The present system was designed to simulate a number of different intonation patterns using a single descriptive apparatus. The systems outlined by 't Hart and Cohen, Lea, and Klatt, by contrast, are designed to simulate neutral declarative intonation only. Thus, these systems do not generate any of the contours just discussed which involve a *L* or *L+H* pitch accent, or a *H* initial boundary tone. Mattingly's system does generate three different phrase final configurations, under the control of diacritics in the input. He has only a single treatment of prenuclear accents, and does not attempt to develop a descriptive apparatus which relates intonation options at the end of the phrase to options elsewhere.

IV. EXPERIMENTAL CONFIRMATION

A number of features of the approach to intonation taken here have been confirmed by experimental work on the production and perception of intonation. Wales and Toner (1980) studied what kinds of ambiguous sentences may be disambiguated using intonation. The three categories of ambiguities they examined were lexical ambiguities, exemplified in sentence (4); deep

structure ambiguities, exemplified in sentence (5); and surface structure phrasing ambiguities, exemplified in sentence (6).

Isn't that what a ruler is for? (4)

Flying planès can be dangerous. (5)

We never fought a bull with real courage. (6)

The only successful disambiguations in their study involved sentences like (6) with two possible surface structure bracketings; homonyms could not be disambiguated using intonation, nor could deep structure ambiguities which did not have a correlate in surface structure bracketing. This result is what the approach taken here would predict since, whereas boundary tones and the distinctive phonetic treatment of the nuclear tone provide ways of marking phrasing, no mechanism is provided for marking the different readings of sentences like (4) or (5). Experiments by Streeter (1978) also confirm that *F0* can be used to disambiguate phrasing.

Nakatani and Schaffer (1978) report experiments on the perception of word boundaries in reiterant speech (speech in which the speaker has been asked to replace some or all syllables of an utterance with the same syllable, here "ma," preserving the prosodic pattern of the model). They found that *F0* is not a cue for word boundary location when the stress contour is fixed. That is, subjects were unable to use *F0* to decide whether the "ma-ma" imitations of the italic words in (7) and (8) represented "mama ma" or "ma mama."

The *noisy dog* kept everyone up all night. (7)

The *bold design* kept everyone's attention. (8)

(Duration differences due to the lengthening of monosyllabic content words could be used with some effectiveness). *F0* was an effective cue for word boundary location only in cases in which it marked a stress pattern which was compatible with only one location of the word boundary. For example, the 1 1 0 stress pattern in (9) would only be possible for "ma mama" and not for "mama ma," given the contextual constraints provided in the experiment.

The *near future* is not yet determined for her. (9)

The present approach predicts these results; *F0* provides a way of marking stress, and given the stress pattern the subject would in some cases be able to infer where the word boundary is. However, for any given stress pattern, the transition between the two targets is insensitive to the location of word boundaries. Thus when the location of the word boundary cannot be inferred from the stress pattern, *F0* provides no further information.

The framework outlined here also carries some predictions about how and when *F0* can be used to mark stress. In the wake of Fry's classic study (Fry, 1958), the impression grew up that *F0* can be viewed as a transducer of stress: The higher the stress, the higher the *F0*. In the framework outlined here, the relation of *F0* to stress is not as direct as this. Rather, a word with a given stress pattern could have any of a number of different *F0* contours, depending on the intonation pattern that was being used. Some *F0* contours would be compatible with more than one stress pattern, while others would permit only one conclusion about the stress pattern. It is only in the second kind of case that *F0* can serve as a cue for stress. In fact, this general picture is supported by experimental work since Fry (1958), as well as by Fry's study itself. Morton and Jassem (1965) report that either lowering or raising the *F0* locally can produce the impression that a syllable is stressed. This means that the perception system does not translate *F0* height directly into stress level. We would expect this result, since either a *L* tone or a *H* tone may be assigned to a stressed syllable. Nakatani and Aston (forthcoming) report that *F0* was not a cue for stress on a noun following a focused adjective. It was mentioned above that no pitch accents are assigned, even to stressed syllables, after the nucleus. Since the focused adjective in Nakatani and Aston's experiment carried the nuclear stress, *F0* could not be used to mark stress on the following noun.

Fry (1958) studied how *F0* and duration influenced judgments of stress on the word "subject," which has initial stress as a noun and final stress as a verb. The interaction of duration with 16 different *F0* contours was examined. He found that some *F0* patterns overrode duration as a cue for stress; that is, for these patterns, subjects gave the same stress judgment more than half the time, regardless of the relative duration of the two syllables. The patterns which best overrode duration as a cue for stress appear to be those for which one intonational analysis would be highly preferred. For example, the two patterns involving a falling *F0* on the first syllable followed by a low *F0* on the second syllable would most readily be interpreted as instances of a nuclear *H* accent on the noun "subject."[4] In contrast, the pattern with a high *F0* on the first syllable and a low and then rising *F0* on the second syllable was judged to be a noun when the first syllable was long and a verb when the second syllable was long. The tabulated results for this contour have 51% noun judgments, suggesting the *F0* contour did not bias stress judgments in either direction. This result does not seem surprising, since the *F0* pattern bears a fair resemblance to either a nonterminal declarative pattern assigned to the noun, or a yes/no question intonation assigned to the verb. The interpretation of results for contours which would not be acceptable in English for either stress pattern is rather unclear. Fry himself suggests that a syllable with *F0* inflection will be perceived as stressed over a level syllable, regardless of the linguistic system. One cannot, however take this to be proved by his experiments. For one, he did not in any way control for effects of the linguistic system on judgments. Secondly, the contours he examined are not a systematic sample of the set of possible contours: for example, he includes results for one contour with an inflected first syllable and a high-level second syllable, but results for four contours with a low-level first syllable and an in-

flected second syllable. Given that the results for two contours he included do not support his conclusion, it seems possible that a different selection of contours would have resulted in quite different averaged results.

V. CONCLUSIONS

The work on English neutral declarative intonation described here was undertaken in a spirit of analysis-by-synthesis. Its outcome was both an approach to analyzing *F0* contours and a working computer program for synthesizing them. The prosodic contribution to the *F0* contour is analyzed as a series of target values within the current *F0* range. Because of the declination effect, the *F0* range narrows and drifts downwards over the course of the phrase. Targets are viewed as belonging to two categories, high and low. When two targets are both high and are sufficiently separated in time, the computer program computes a sagging *F0* contour in between. Otherwise, the *F0* contour between the two targets is a monotonic curve. Segmental effects on *F0*, which are not modeled in the computer program, introduce local perturbations on the prosodic *F0* contour which can be seen in *F0* contours of natural utterances.

We can identify several ways in which the approach to synthesizing intonation taken here proved successful. It permitted a very good synthesis of neutral declarative intonation from quite a sparse input representation. The description of intonation which was used to generate the synthetic *F0* contours conforms to experimental results on how *F0* may be used to mark stress and phrasing. Unlike the frameworks used in other synthesis programs, it extends in a straightforward way to permit the synthesis of other intonation patterns. For this reason, the program makes a useful tool for research into the full system of English intonation.

Work with the program pointed up some gaps in our knowledge of the phenomenology of intonation. The timing of the tones at the end of the phrase needs to be better understood. The range of target values for *L* was set quite arbitrarily here, since there are few data from natural speech on this point. The quality of the synthetic *F0* contours would also be improved if the program modeled the segmental effects on *F0*. In order to do this accurately, we need more data on how segmental effects interact with stress and intonation pattern in running speech.

ACKNOWLEDGMENTS

I would like to thank Mark Liberman and Osamu Fujimura for their help and thoughtful criticisms.

[1]Using linear predictor coded speech makes it possible to change the *F0* contour of an utterance while maintaining essentially the original segmental characteristics. See Atal and Hanauer (1971) for details.

[2]The *F0* synthesis program was designed to be used with the dyad synthesis-by-rule system described in Olive and Liberman (1979). It is also being used with the demisyllable synthesis-by-rule system described in Lovins *et al.* (1979) and Browman (1979).

[3]This constraint on relative peak height is intended to apply only in the case of the neutral declarative intonation pattern, which has a H on each accented syllable. There is a different intonation pattern, referred to in Crystal (1969) and O'Connor and Arnold (1961) as the "stepping head," in which the targets form a descending staircase. In this pattern the highest peak falls on the first stressed syllable rather than on the most stressed one. A theoretical analysis of this contour in the two-tone framework used here is presented in Pierrehumbert (1980) together with relevant phonetic data.

[4]Two alternative interpretations, discussed in Pierrehumbert (1980), are unusual intonation patterns which would not seem natural without a discourse context which strongly motivates them.

Allen, J., Hunnicutt, S., Carlson, R., and Granstrom, B. (1979). "MITalk-79: The MIT Text-to-Speech System," *in Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 507–510.

Ashby, M. (1978). "A Study of Two English Nuclear Tones," Lang. Speech 21, 326–336.

Atal, B. S., and Hanauer, S. L. (1971). "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am. 50, 637–655.

Bolinger, D. (1951). "Intonation: Levels versus Configurations," Word 7, 199–210.

Browman, C. P. (1979). "Lingua: A Language Interpreter Used for Demisyllable Synthesis," J. Acoust. Soc. Am. Suppl. 1,66, S23.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective* (CWK Gleerup, LiberLaromedel, Lund).

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

Clark, M. (1978). "A Dynamic Treatment of Tone, with Special Attention to the Tonal System of Igbo," Ph. D. dissertation, University of Massachusetts at Amherst (distributed by Indiana University Linguistics Club, Bloomington).

Cooper, W., and Sorensen, J. (1977). "Fundamental Frequency Contours at Syntactic Boundaries," J. Acoust. Soc. Am. 62, 683–692.

Crystal, D. (1969). *Prosodic Systems and Intonation in English* (Cambridge U. P., London).

Fry, D. B. (1958). "Experiments in the Perception of Stress," Lang. Speech 1, 125–152.

Fujisaki, H., and Nagashima, S. (1969). "A Model for the Synthesis of Pitch Contours of Connected Speech," Annual Report of the Engineering Research Institute, Tokyo, 28, 53–60.

Klatt, D. (1979). "Synthesis by Rule of Segmental Durations in English Sentences," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 287–301.

Ladd, D. R. (1978). "The Structure of Intonational Meaning," Ph. D. dissertation, Cornell University. (Indiana U. P., Bloomington, 1980).

Lea, W. A. (1972). "Intonation Cues to the Constituent Structure and Phonemics of Spoken English," Ph. D. dissertation, Purdue University.

Lea, W. A. (1973). "Segmental and Suprasegmental Influences on Fundamental Frequency Contours," in *Consonant Types and Tones*, edited by L. Hyman (Southern California Occasional Papers in Linguistics, Los Angeles), pp. 15–70.

Lehiste, I., and Peterson, G. E. (1961). "Some Basic Considerations in the Analysis of Intonation," J. Acoust. Soc. Am. 33, 419–425.

Liberman, M. (1975). "The Intonation System of English,"

Ph. D. dissertation, MIT (Garland, New York, 1979).
Liberman, M. and Pierrehumbert, J. (1979). "A Metric for the Height of Certain Pitch Peaks in English," J. Acoust. Soc. Am. Suppl. 1, 66, S64.
Liberman M., and Prince A. (1977). "On Stress and Linguistic Rhythm," Linquistic Inquiry 8, 249–336.
Liberman, M. and Sag, I. (1974). "Prosodic Form and Discourse Function," in *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, edited by M. LaGaly, R. Fox, and A. Bruck (Chicago Linguistic Society, Chicago), pp. 416–427.
Lovins, J. B., Macchi, M. J., and Fujimura, O. (1979). "A Demisyllable Inventory for Speech Synthesis," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 519–522.
Maeda, S. (1976). "A Characterization of American English Intonation," Ph. D. dissertation, MIT.
Mattingly, I. (1966). "Synthesis By Rule of Prosodic Features," Lang. Speech 9, 1–13.
Mattingly, I. (1968). "Synthesis by Rule of General American English," Supplement to Status Report on Speech Research (Haskins Laboratories, New Haven).
Morton, J., and Jassem, W. (1965). "Acoustic Correlates of Stress," Lang. Speech 8, 159–181.
Nakatani, L., and Aston, C. (forthcoming). "Perceiving the Stress Pattern of Words in Sentences," Phonetica (to be published).
Nakatani, L., and Schaffer, J. (1978). "Hearing 'Words' Without Words: Prosodic Cues for Word Perception," J. Acoust. Soc. Am. 63, 234–244.
Nooteboom, S. G., Brokx, J. P. L., and de Rooi, J. J. (1976). "Contributions of Prosody to Speech Perception," *Studies in Language Perception*, Proceedings of the Symposium on Language Perception, International Congress of Psychology, Paris.
O'Connor, J. O., and Arnold, G. F. (1961). *Intonation of Colloquial English* (Longmans Green, London).
Öhman, S. (1967). "Word and Sentence Intonation: A Quantitative Model," Speech Transmission Laboratory Quarterly Progress and Status Report 2–3, 20–54.
Olive, J. (1974). "Speech Synthesis by Rule," Speech Communication Seminar, Stockholm, August 1–3.
Olive, J., and Liberman, M. (1979). "A Set of Concatenative Units for Speech Synthesis," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 515–518.
Olive, J., and Nakatani, L. (1974). "Rule-Synthesis of Speech by Word Concatenation: A First Step," J. Acoust. Soc. Am. 55, 660–666.
O'Shaughnessy, D. (1976). "Modelling Fundamental Frequency and its Relationship to Syntax, Semantics, and Phonetics," Ph. D. dissertation, MIT.
Pierrehumbert, J. (1979a). "Intonation Synthesis Based on Metrical Grids," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 523–526.
Pierrehumbert, J. (1979b). "The Perception of Fundamental Frequency Declination," J. Acoust. Soc. Am. 66, 363–369.
Pierrehumbert, J. (1980). "The Phonology and Phonetics of English Intonation," Ph. D. dissertation, MIT (forthcoming from MIT Press).
Pike, K. L. (1945). *The Intonation of American English* (University of Michigan, Ann Arbor).
Sag, I., and Liberman, M. (1975). "The Intonational Disambiguation of Indirect Speech Acts," in *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*, edited by R. Grossman, L. J. San, and T. Vance (Chicago Linguistic Society, Chicago), pp. 487–497.
Sorensen, J., and Cooper, W. (1980). "Syntactic Coding of Fundamental Frequency in Speech Production," *Perception and Production of Fluent Speech*, edited by R. A. Cole (Lawrence Erlbaum, Hillsdale), pp. 399–440.
Streeter, L. (1978). "Acoustic Determinants of Phrase Boundary Perception," J. Acoust. Soc. Am. 64, 1582–1592.
't Hart, J., and Cohen, A. (1973). "Intonation by Rule: A Perceptual Quest," J. Phonetics 1, 309–327.
Trager, G. L., and Smith, H. L. (1951). *Outline of English Structure* (Battenburg, Norman, OK).
Vaissière, J. (1971). "Contribution à la Synthèse par Regles du Francais," Doctoral dissertation, Universitè des Langues et Lettres de Grenoble.
Vanderslice, R., and Ladefoged, P. (1972). "Binary Suprasegmental Features and Transformational Word-Accentuation Rules," Language 48 (4), 819–839.
Vives, R., le Corre, G., Mercier, G., and Vaissière, J. (1977). "Utilisation, pour la Reconnaissance de la Parole Continue, de Marouers Prosodiques Extrais de la Frequence du Fondamental," Recherches Acoustiques 4, 237–252.
Wales, R., and Toner, H. (1980). "Intonation and Ambiguity," *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, edited by W. E. Cooper and E. C. Walker (Lawrence Erlbaum, Hillsdale) pp. 135–158.