ultimately erasing) effects in segmental phonology and prevents the vowels from undergoing certain tonal processes.

The thesis of this article may be regarded as an illustration of the view that unified phonetic explication is possible for phonological processes that are substantively rather than formally related. In this case a range of apparently unrelated segmental and tonal phenomena have been shown to converge on the sonority hierarchy, which is independently motivated by phonetic considerations of defining possible syllabic structure.

Notes

This paper is based on a preliminary version read at the monthly meeting of the Circle of Spoken Language Studies (CSLS) at the University of Tsukuba on December 1, 1981. I am especially grateful to Minoru Yasui and the members of the CSLS for providing me with many interesting comments and suggestions. I am also grateful to N. Teele for reading the draft. None of them are, however, responsible for the remaining errors or insufficiencies.

[1] There seems to be some dialectal fluctuation is these cases. Thus, in my own dialect these accented high vowels also undergo devoicing, which means that in my dialect high vowels are all devoiced in the environment under consideration.

[2] For some discussion of this, see Haraguchi (1977).

[3] For a more detailed discussion of the Flop rule and other phenomena in the Takamatsu dialect, see Haraguchi (1977).

[4] I am grateful to Minoru Yasui for pointing out this notion to me.

[5] I assume, in addition to the sonority hierarchy, that H tone is more sonorous than L tone and that stress also affects sonority. See Liberman and Pierrehumbert (this volume) for relevant discussion.

Chapter 10

Intonational Invariance under Changes in Pitch Range and Length

Mark Liberman
Janet Pierrehumbert

1. Introduction

The central question of phonological theory is the nature of phonological substance (primitive entities and relations) and of its connections with words and sounds. Taking the realization of abstract intonational categories as a case in point, we will argue that the links between phonological substance and sound are more complex and more consequential than usually assumed. An appreciation of this point will bring better linguistic descriptions in terms of simpler linguistic theories; or so we hope.

The work reported here arises from our interest in the interaction of four factors in the description of English intonation: tune, prominence, declination, and pitch range. The nature of these four factors is illustrated in figures 1 through 8 and discussed in the next four subsections. Briefly, in our usage *tune* refers to intonation contour type, *prominence* to local degree of stress or emphasis, *declination* to a downward trend in pitch across a phrase and *pitch range* to a global, or at least phrase-sized, choice of pitch-scaling parameters.

The status of these factors in past linguistic descriptions of intonation differs. Tune falls most clearly within the scope of such descriptions, while the other factors are sometimes set aside as phonetic or paralinguistic. Our experience in working on intonation, however, has led us to believe that (as a practical matter) no one of these factors can be understood without an understanding of the others. As a result, a model of the entire system is needed in order to decide linguistic issues in the study of intonation, such as the number and type of tonal categories, how they are organized phonologically, and how their phonological organization is related to syntactic and semantic structure.

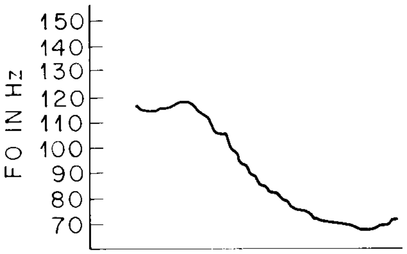


Figure 1
An F0 contour for the monosyllable *Anne*, produced with a typical declarative intonation.

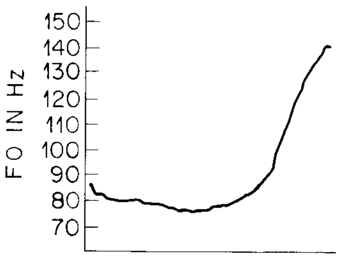


Figure 2
An F0 contour for *Anne*, produced with a typical interrogative intonation.

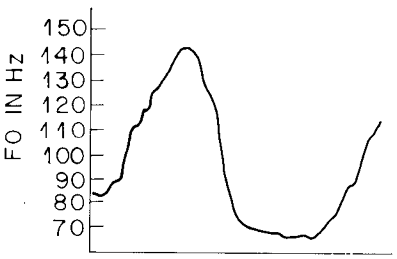


Figure 3
An F0 contour for *Anne*, produced with a rise-fall-rise intonation pattern that can convey incredulity.

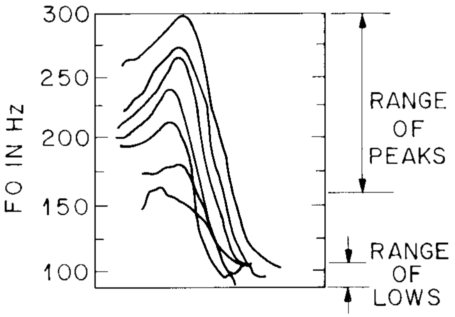


Figure 4
A declarative intonation pattern produced on *Anne* with seven different amounts of overall emphasis. The peak values vary widely, while the terminal low values remain essentially constant.

1.1 Tune

Figures 1 through 4 show fundamental frequency (F0) contours for three different tunes produced on the same monosyllable, *Anne*. F0 is the main physical counterpart of our impressions of pitch, and the F0 contours shown here were obtained by computer analysis of a digitized speech waveform. The contour shown in figure 1 is a typical declarative intonation pattern. It has a peak followed by a fall to the bottom of the speaker's pitch range. The F0 contour in figure 2 has a low value followed by a rise; this pattern is often used to ask a question. The rise-fall-rise pattern shown in figure 3 can be used to convey incredulity.

The linguistics literature on intonation contains many proposals for representing English tunes phonologically. In the first part of this paper, our assumptions about the character of this representation will be minimal. We will assume that tunes can be decomposed into sequences of elements that are aligned with the text. These elements include pitch accents, which mark some (but not necessarily all) stressed syllables, and additional tonal features associated with the end of the intonation phrase.

The two types of pitch accent that we investigated experimentally are shown in figures 5 and 6. For now, we will refer to the accent type shown in figure 5 as a *peak accent*, since it is realized as a peak on the stressed syllable. The accent circled in figure 6 will be called a *step accent*. The step accent has a relatively lower level on its stressed syllable and a relatively higher level just before. It is important to note that the lower level of one step accent becomes the higher level of a succeeding step accent, so that a sequence of step accents forms a descending staircase. In section 6, we will

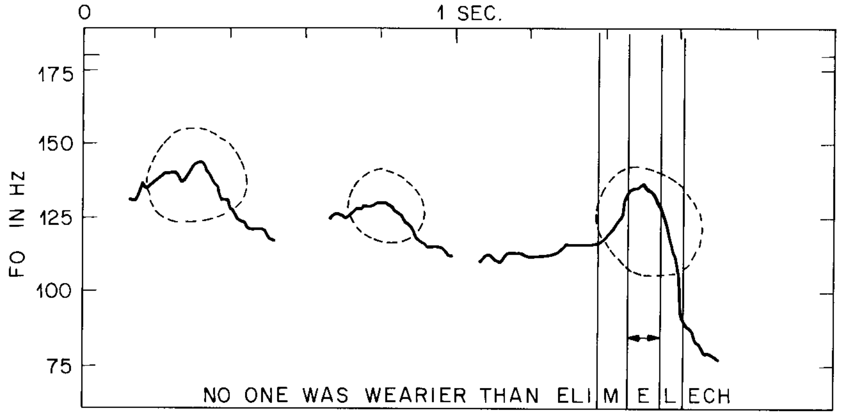


Figure 5
An intonation phrase containing three peak accents, indicated with dashed circles. *Wearier* has a lower peak than *no* and *Elimelech* because it is less prominent.

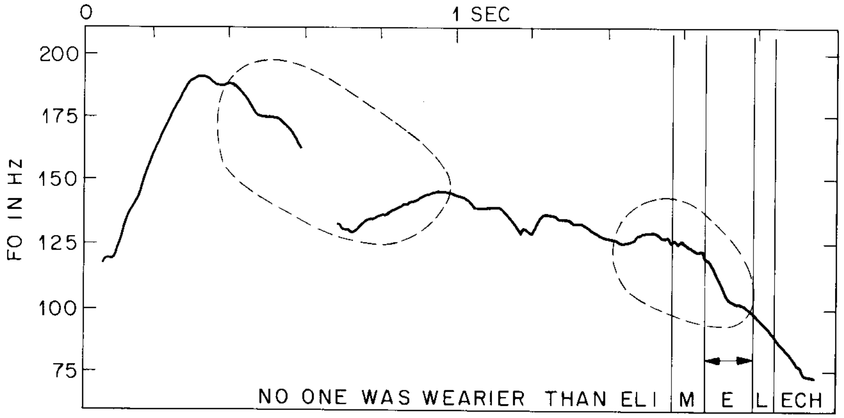


Figure 6
An intonation phrase containing two step accents, indicated with dashed circles. Compare the configuration on the stressed vowel in *Elimelech*, indicated with a double arrow, to the corresponding region in figure 5.

relate the results of our experiments to the theory of English tonology developed in Pierrehumbert (1980). In this account, tunes are structured sequences of Low and High tones. We will suggest that our approach to modeling the realization of peak and step accents can be extended to cover the qualitative behavior of L and H tones in other contexts. In addition, some of our data tend to support the idea that tunes are decomposed into target levels, as proposed here, rather than into pitch changes as other authors have suggested.

1.2 Prominence

A given tune can be produced as many systematically different F0 contours, even by the same speaker. One factor controlling such variation is the prominence of the material with which the tune is associated. Figure 4 shows the declarative intonation pattern of figure 1 produced with seven different degrees of emphasis: more emphasis results in a higher peak. Figure 5 shows that pitch accents within one phrasal tune can separately reflect prominence in this way. *Wearier* is less strongly stressed than *no* and *Elimelech*, and the peak accent associated with it is accordingly lower.

In contrast to tune differences, which are qualitative, emphasis or prominence differences appear to be quantitative. That is, the underlying parameter is continuously variable. We used only seven F0 contours to make up figure 4, for the sake of graphical legibility; it would have been possible to supply many intermediate cases.

1.3 Declination

In general, a given pitch accent under a given amount of emphasis will give rise to different F0 values in different parts of a phrase. A number of researchers[1] have reported that the range of F0 values employed is narrower and lower at the end of a phrase than at the beginning; this phenomenon is given the name *declination*. Figures 7 and 8 illustrate two observations of the kind that have motivated such reports. Figure 7 is the F0 contour of a complex declarative sentence. Peaks in the F0 contour bob up and down, but there seems to be a general downward trend. Figure 8 shows a large number of declarative F0 contours that have been time-normalized and averaged. The utterances used were the first 43 phrases produced by a radio announcer in an extemporized monologue. A phrase boundary was posited wherever there was a nonhesitation pause or a clear sentence boundary. Again, there seems to be a downward trend. It is important to note, however, that the shape in this figure could be due in part to the time-normalization or to the speaker's use of step accents. In one of the experi-

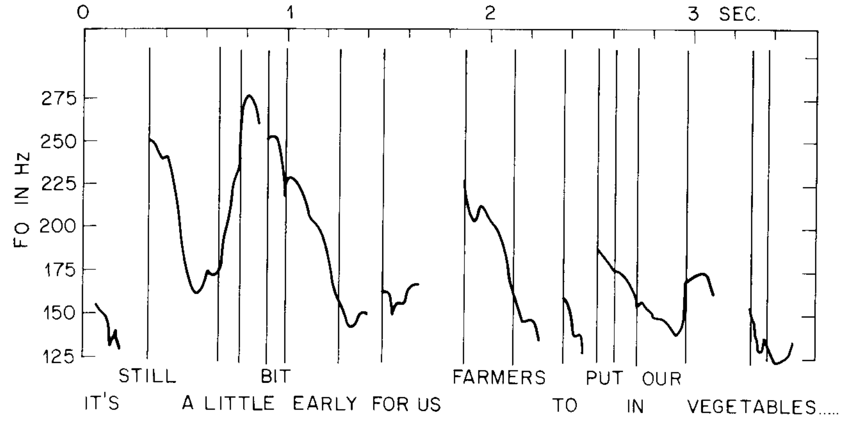


Figure 7
The F0 contour of an utterance produced during a radio talk show. Overall downtrends like the one observed in this contour have motivated reports of declination.

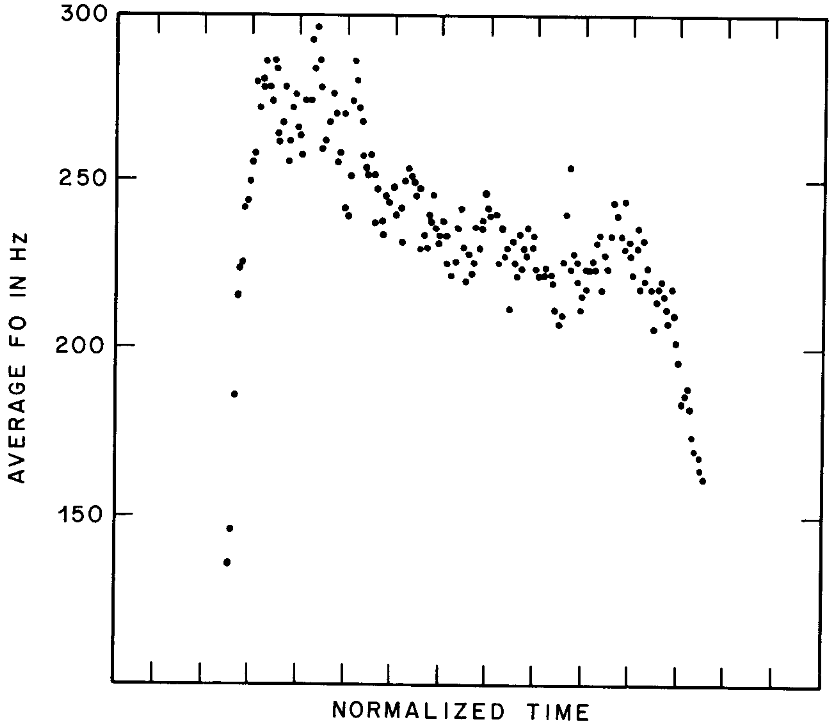


Figure 8
The F0 contours for the first 43 phrases produced by the same radio announcer, time-normalized and averaged. There is a gradual downtrend in the values.

ments described below, we have tried to determine the time course of declination more reliably.

One consequence of declination is that two consecutive peak accents with the same peak F0 value need not, in general, count as having the same prominence. In appropriate experiments, listeners normalize for the declination effect in computing relative prominence, so that the second of two equal accents in general sounds higher. For two accents to sound equally prominent, the second must in general have a lower F0 value (Pierrehumbert (1979)).

1.4 Overall Pitch Range

There is a fourth dimension of intonational performance, it seems, besides those we have mentioned so far. A speaker may, for instance, "speak up" in order to be heard at a distance or through noise, or pitch his or her voice higher in order to seem small or cute. Both of these modes of production affect the overall scaling of the F0 contour in a way that cuts across variations arising from tune choice, emphasis, and phrasal position. Their effects on overall scaling are somewhat different, and no doubt other paralinguistically motivated types of variation can be found, with yet different effects on overall scaling. Here, we will be concerned with effects on overall scaling due to "speaking up." Investigation of how the effect of "speaking up" interacts with the other factors determining the F0 contour will suggest a way of modeling the overall scaling function.

1.5 Degrees of Freedom

The four terms we have just introduced are pretheoretical ones. If we hope to develop a good description of intonation, it will not do to simply throw these terms (along with any others we might think of) into a sort of stew-pot of descriptive categories. For one thing, it seems difficult to get agreement on a set of descriptive categories. Even if we negotiated a consensus, it is quite likely that we would get the primitives wrong—intuitively derived descriptive categories are commonly found to be complex combinations of initially unintuitive basic entities.

Furthermore, we must be careful that our descriptive system preserves the ability to characterize individual intonation patterns without excessive ambiguity. For this reason, an important issue in any discussion of intonational representations is the number of degrees of freedom in the system as a whole. This issue was first raised by Bolinger (1958) in his discussion of pitch range and the phonological representation of tune. As he noted, English tonal specifications are sparse relative to the rate with which it is possible to vary pitch expressively (i.e., through changes in quantitative variables like emphasis and pitch range). As a result, a description of English intonation that has four tone levels[2] seems to have too many degrees of freedom. For example, it is hard to tell whether an F0 contour that rises and then falls represents 4–1–4 in one pitch range, 4–2–4 in a larger pitch range, or 4–3–4 in a still larger one. At least, the burden of proof is on the proponents of four-tone systems to show how the claimed information content can be conveyed.

Reducing the tonal inventory to two tones lightens the burden considerably. However, declination, prominence, and pitch range variation reintroduce the problem in a different guise. Consider only the interaction of declination and prominence. If the pattern of declination were completely predictable, then the listener or linguist could easily factor it out to arrive at an analysis of the tune and prominence relations. On the other hand, if the declination pattern were unpredictable, perhaps because it varied to express some form of meaning, then it would have to be recovered from the speech signal along with the tune and prominence relations. On this second assumption, it would be a challenge to explain how these variables could be recovered from the information available to the listener. The same issue arises again when the interaction of emphasis and overall pitch range is considered.

1.6 Methods: An Experimental Approach to Intonational Description

The problems we have just discussed arise from the apparent existence of several sources of quantitative variation in the realization of English intonation patterns. The true number and nature of these dimensions of variation is initially unknown, as is the true character of the patterns they modulate. We have tried to use this same variation as a tool to uncover, at least partly, the underlying intonation system.

In phonological research, variations in the form of a stem under inflection are used as evidence about its underlying form; the underlying form of an ending is similarly uncovered by suffixing it to different stems. In our experiments on intonation, we have adapted this approach in order to apply it to quantitative data. Factors that affect the phonetic realization of an intonational parameter or category are explicitly and systematically varied. The patterns apparent in the resulting data are then used in constructing a theory.

In order to keep the task manageable, the sources of variation in each experiment must be limited. This is partly because large amounts of data must be collected to compensate for uncertainties of production and measurement, but it primarily reflects the need to limit the conceptual scale of each experiment, so that the search for structure can be thorough and rigorous. Broader coverage can then be achieved by applying the same models, with the same values for relevant free parameters, to the results of several experiments at once.

1.7 Models of Intonation and Their Linguistic Interest

In the body of this paper, we will describe two experiments and the models we arrived at for describing the data from them. By a model, we mean an explicit system of rules for predicting measurements on the basis of linguistic and paralinguistic properties of assumed descriptions. Constructing such a model brings out the structure of the system under study. While the models we will present are particular to the kinds of intonational variation in our experiments, we believe that the results are sufficient to suggest the general properties of the system through which tune, prominence, declination, and pitch range interact in English intonation.

The models we will present might be viewed as models of phonetic realization. However, they also bear on important issues in phonology, syntax, and semantics. Some of these issues are descriptive ones, others are methodological, and still others are theoretical.

1.7.1 Some Descriptive Issues Connections to certain issues of traditional concern in linguistics are inherent in any description of intonational phenomena. Intonational domains, or structural units, are often taken to be related in some way to various phonological, syntactic, and semantic units, and the meaning of intonational categories is often seen as crucial to other aspects of semantic interpretation. Thus, observations about stress and intonational phrasing have played a role in arguments about syntactic structure (e.g., Bresnan (1971), Williams (1974)), and discussions of the relation of semantic interpretation to syntactic structure have made essential reference to intonational descriptions. In particular, focus, presupposition, and scope of negation and quantification are said to be related to stress, phrasing, and tune choice (see Chomsky (1971), Jackendoff (1972), Katz (1972), Carlson (1982)). We will not comment on these subjects here, but wish only to note the importance of describing and categorizing the relevant intonation patterns correctly.

There has been a lively debate about the nature of tonal features, with some favoring "static" features like "high" and "low" while others favor "dynamic" features like "rise" and "fall." In section 5.3, we will argue that our observations strike a blow in favor of static features.

1.7.2 Methodological and Theoretical Issues Our methods, which are somewhat unusual, combine the phonologist's traditional concern for relations among abstract representations with the phonetician's interest in accounting for the details of actual speech. Our experience with these hybrid methods suggests that the correct "division of labor" between abstract phonological descriptions and the process of phonetic interpretation is not easy to discover. This point, applied to the subject matter of segmental phonology, will lead us to raise some pointed questions, in section 7, about the correct treatment of allophonic variation. A reasonable answer to these questions would force most "postlexical" phenomena (in the sense of Mohanan (1982) and Kiparsky (1982)) to be treated as facts about the phonetic realization of phonological representations, rather than as modifications of phonological representations themselves.

In contrasting our models with some other proposals in the literature, we will be led to suggest certain restrictions on the use of hierarchical representations in phonology. Such representations were originally introduced in order to eliminate nonlocal dependencies from phonological rules, by making available a limited set of adjacency relations beyond those inherent in a phonemic string, and by permitting features to be defined over a limited set of domains larger than the phoneme.

1.8 A Sketch of Our Conclusions

First, the preconditions necessary for our method to be a useful one appear to exist. Subjects are able to produce the kinds of variation we require; the results are quite lawful and are similar across individuals.

Second, a class of successful models can be found. In these models, the contributions of pitch range and phrasal position are combined in a simple way with the realization of phonological tone patterns and relative prominence relations. Thus, complex patterns of data emerge from the interaction of several simple sources of variation.

Third, we find no clear evidence for phrase-level planning of F0 implementation. The major factors shaping F0 contours appear to be local ones. Potential long-distance effects are erratic and small at best.[3] All of our model's computations are made on pairs of adjacent elements, and the parameters of the transform need not be set differently for different phrase lengths. Significant phrasal position effects seem to be limited to a lowering of final pitch accents and therefore would fall within the scope of the two-pitch-accent window.

The phenomena collectively categorized as "declination" are, in our theory, explained by a combination of the final lowering effect, the frequent usage of stepping accents, and perhaps the statistics of relative prominence. We have no reason to think that the final lowering effect is varied for expressive purposes. As a result, its separation from tune and prominence is fairly simple.

On the other hand, local prominence and overall pitch range do seem to represent independently variable quantities, and in some cases their separation is therefore problematic. In the experiments to be described, under the models we have used, the overall pitch range and the prominence of the initial pitch accent co-vary in a way that permits the full description to be uniquely recovered from our measurements.

2. Description of Experiments: Materials and Procedures

2.1 Materials

The two experiments we will discuss investigated how particular intonation patterns are scaled under changes in overall pitch range. In each experiment, one additional characteristic of the intonation pattern was also varied.

The first experiment investigated the two intonation patterns shown in figures 9 and 10.[4] The pattern in figure 9 is produced as the response in the following dialogue:

(1)
Question: What about Manny? Who came with him?
Answer: Anna came with Manny.

Here, *Anna* is really the answer to the question, and *Manny* counts as background information. In the pattern shown in figure 10, the order of the answer and the background information is reversed. This pattern might be produced as the response in the following dialogue:

(2)
Question: What about Anna? Who did she come with?
Answer: Anna came with Manny.

In neither case is the cited pattern the only way to produce the answer sentence, given the cited question, and subjects must be instructed by example in order to ensure that the cited pattern will be the one chosen. However, the cited patterns and priming questions may not be interchanged: the pattern in figure 10 is not an appropriate answer to the question in (1), nor is the pattern in figure 9 possible as an answer to the question in (2).

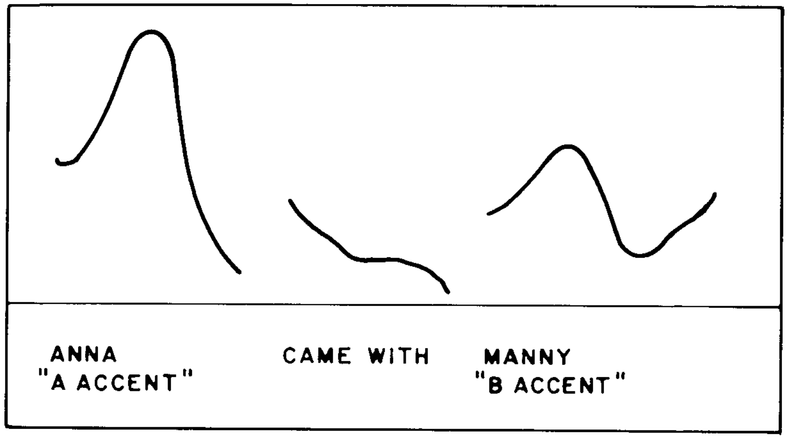


Figure 9
An F0 contour for *Anna came with Manny*, produced as a response to *What about Manny? Who came with him?*

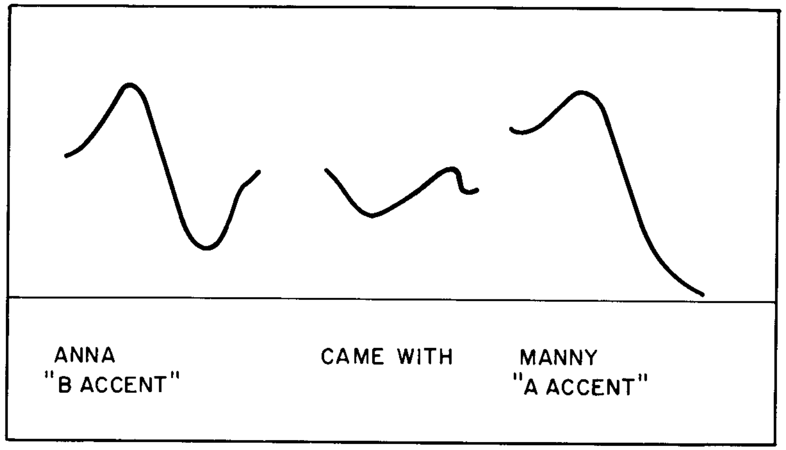


Figure 10
An F0 contour for *Anna came with Manny*, produced as a response to *What about Anna? Who did she come with?*

In both of these patterns, the main answer is more prominent than the background information and is accordingly produced with a higher peak F0 value. However, in figure 9 the peak on the answer is much higher than the peak on the background phrase, while in figure 10 the peak on the answer is only slightly higher. Our hypothesis is that the underlying prominence relation is the same in both cases, and that the difference arises because the final lowering effect adds to the result of the prominence difference in the first case, but partly offsets it in the second case. Under an appropriate transform, the ratio of the "answer" and "background" peak F0 values will be constant, regardless of their order.

The tunes as well as the prominence differences in the intonation patterns in figures 9 and 10 will be important in developing our model. We note that both patterns have two intonational phrases; that is, each of the pitch accents in the sentence is the nuclear pitch accent of its own phrase. Thus, the patterns investigated contrast with the single-phrase patterns that would also be appropriate in the same contexts, shown in figures 11 and 12. In both of the patterns in figures 9 and 10, the peak on the main answer is followed by a fall to the bottom of the speaker's range. The fall following the peak on the background phrase stops somewhat short of this and is followed by a rise at the very end of the phrase. We will refer to the entire complex of intonational properties that goes with the answer as the *A configuration* and the complex that goes with the background phrase as the *B configuration.*[5] Thus, the pattern in figure 9 is an AB pattern, and the one in figure 10 is a BA pattern.

In the first experiment, the pitch range instruction was varied in 10 steps, and six to eight repetitions of each pattern in each pitch range were recorded. In both of the experiments to be described, "degree of overall emphasis or excitement" was the term used in the subjects' instructions, and the kind of variation desired was illustrated by example. This instruction produces simultaneous variation in pitch range, amplitude, and rate, the more "emphatic" or "excited" versions being higher pitched, louder, and slower. The intention was not to produce "pure" variation in a single measurement or on a single psychological scale, but rather to permit the subjects to produce variation in as natural a way as possible.

The second experiment involved downstepping contours (staircases of stepping accents) on lists of berry names, of the sort illustrated in figure 13. In an example such as this, with a large number of steps, it is clear that each step is smaller than the one before; the overall impression is of an exponential curve. This is just what we would expect if the step size were a constant fraction of the preceding level. Anderson (1978) proposes that downstep in

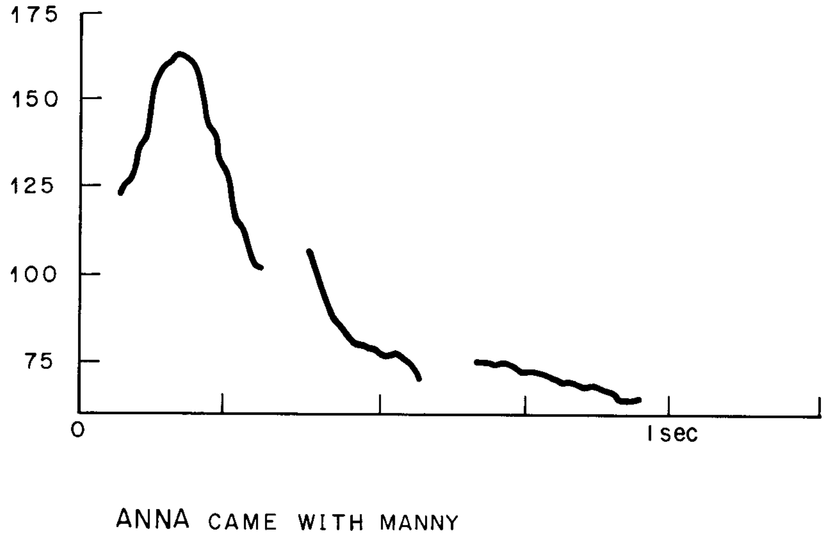


Figure 11
Single-phrase answer to the same question asked in figure 9.

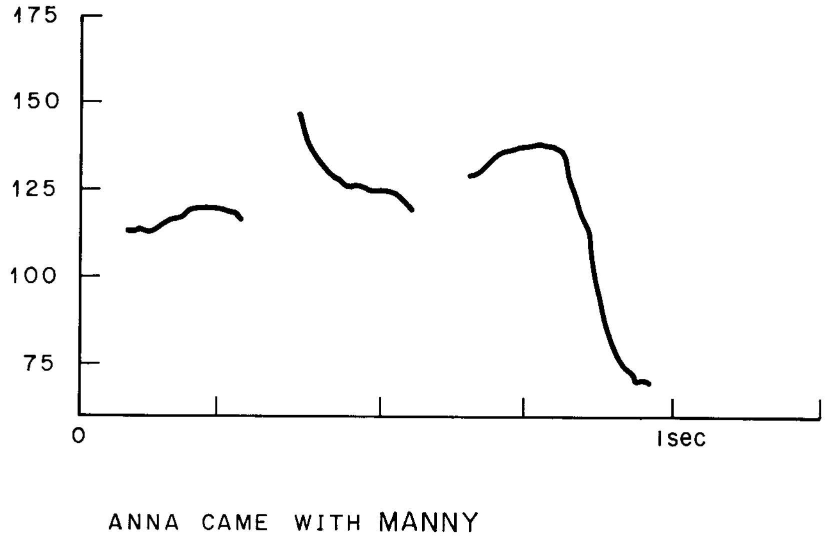


Figure 12
Single-phrase answer to the same question asked in figure 10. These patterns (along with many other possible contours) were not studied in the first experiment.

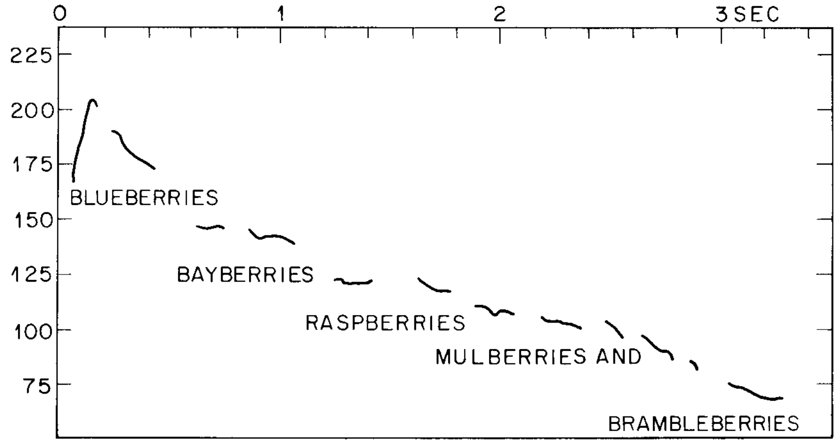


Figure 13
An F0 contour for the berry list *Blueberries, bayberries, raspberries, mulberries, and brambleberries*, produced with a sequence of step accents. Each step is smaller than the one before, so that the step levels appear to trace out an exponential decay.

African tone languages be treated as an abstract exponential decay generated by a constant ratio rule. In the rest of this paper, we will show that under an appropriate transform, successive F0 levels in English stepping contours can be predicted quite accurately by a theory of this type.

The list length varied from two to five, and the pitch range was varied in three steps. The berry names were drawn from the set *bilberries*, *raspberries*, *bayberries*, *blueberries*, and *mulberries*. Twenty different lists of each length were used. For list length two, we used all possible ordered pairs of unlike berries. For each longer length, the set of lists had all berries equally represented in all serial positions, and all transitional probabilities equal. Permuted lists were used as the materials for two reasons. First, all words in such a list tend to have equal importance, so that semantic or pragmatic effects on relative prominence are minimized. Second, by averaging peak F0 values in a given list position across lists of a given length, the effects of speech segments on measured F0 could be removed.[6] The five berry names were used in preference to numbers or letters because they are voiced throughout, and because two syllables separate the accented syllables from each other and from the end of the phrase. These two properties facilitate measurement and analysis.

Studying the stepping intonation pattern allows us to address several questions that cannot be addressed using only data from the first experiment. The presence of up to five list items lets us sample possible serial position effects in more than two places. The differing length of the lists allows us to look for evidence of preplanning. Also, the second experiment, unlike the first, shows effects inside a single intonational phrase.

2.2 Procedures

Recordings were made in a sound-treated booth. For the first experiment, there were four subjects, of which two were the authors and two were Bell Labs summer employees. Two were women and two were men. Three of these subjects also participated in the second experiment. In both experiments, each subject was given a stack of note cards with the materials to be read. In the first experiment, each card listed:

(a) The question associated with the desired intonation, as in (1) and (2).

(b) The response, *Anna came with Manny*.

(c) A number from 1 to 10, indicating the degree of "overall emphasis" to be used in producing the response.

Subjects read both the question and the response. Cards were randomized in blocks of 20 consisting of all combinations of pitch range and response type.[7] For the first subject, six such blocks were recorded, and for subsequent subjects, eight blocks. For subjects other than the authors, the desired intonation patterns were demonstrated by example before the experiment, and the ability of the subjects to produce them naturally was checked.

The meaning of the "overall emphasis" scale was also demonstrated, using a one-word utterance: "1" was used for a phlegmatic mumble, while "10" was used for a forceful shout. Varying the intonation pattern and overall emphasis orthogonally, as the experiment required, did not seem very difficult. As the data plots in figures 14 through 17 indicate, a wide variety of pitch ranges was elicited. We should also note that a corresponding range of amplitudes was also produced and that the most emphatic renditions were up to twice as long as the least emphatic.

In the second experiment, each card read by the subjects had:

(a) A list of two to five berry names.

(b) A number from 1 to 3 indicating the degree of overall emphasis to be used in reading the list.

There were 20 distinct lists of each length, four lengths, and three pitch ranges. The 240 resulting cards were shuffled in blocks of 12. Each block contained all four lengths and three pitch ranges, with no repetitions of the same berry name list in any block. The resulting ordering was checked for

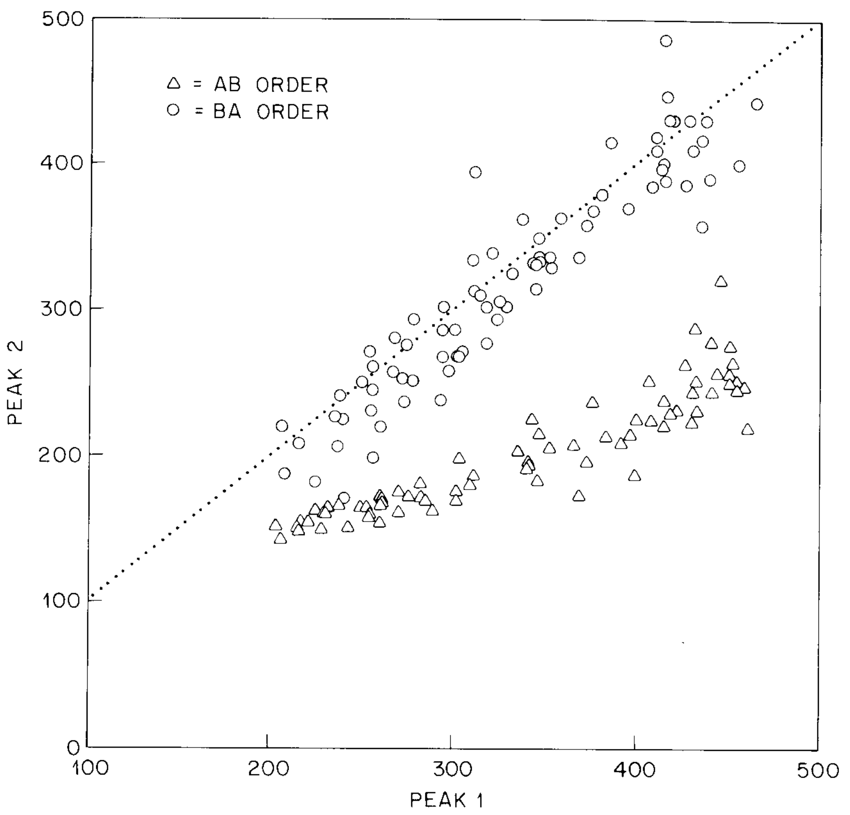


Figure 14
Answer-Background peak data for subject KXG

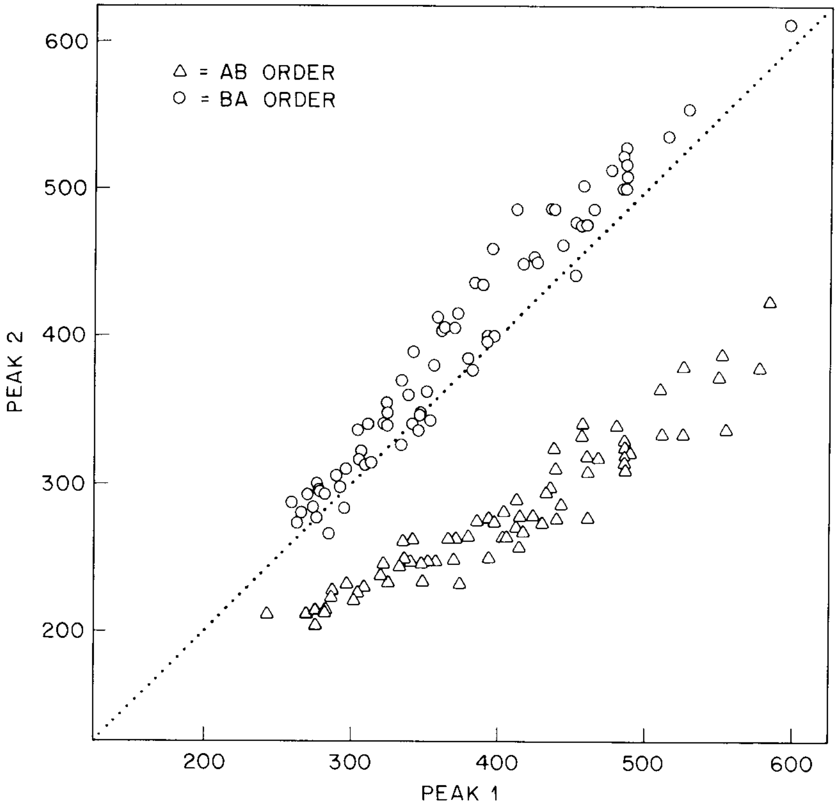


Figure 15
Answer-Background peak data for subject JPB

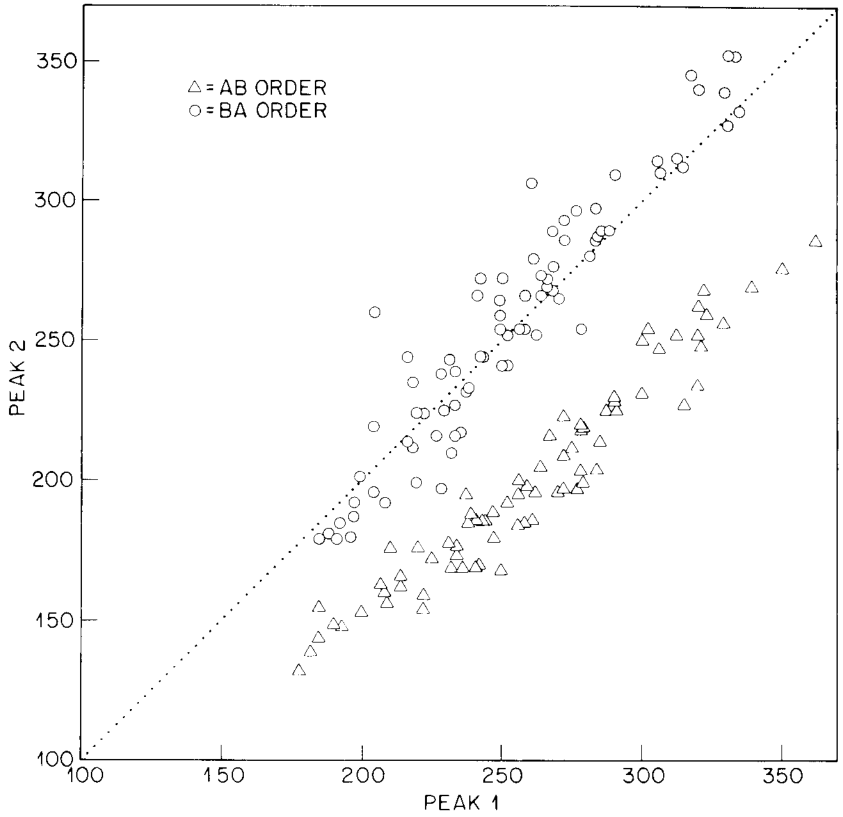


Figure 16
Answer-Background peak data for subject DWS

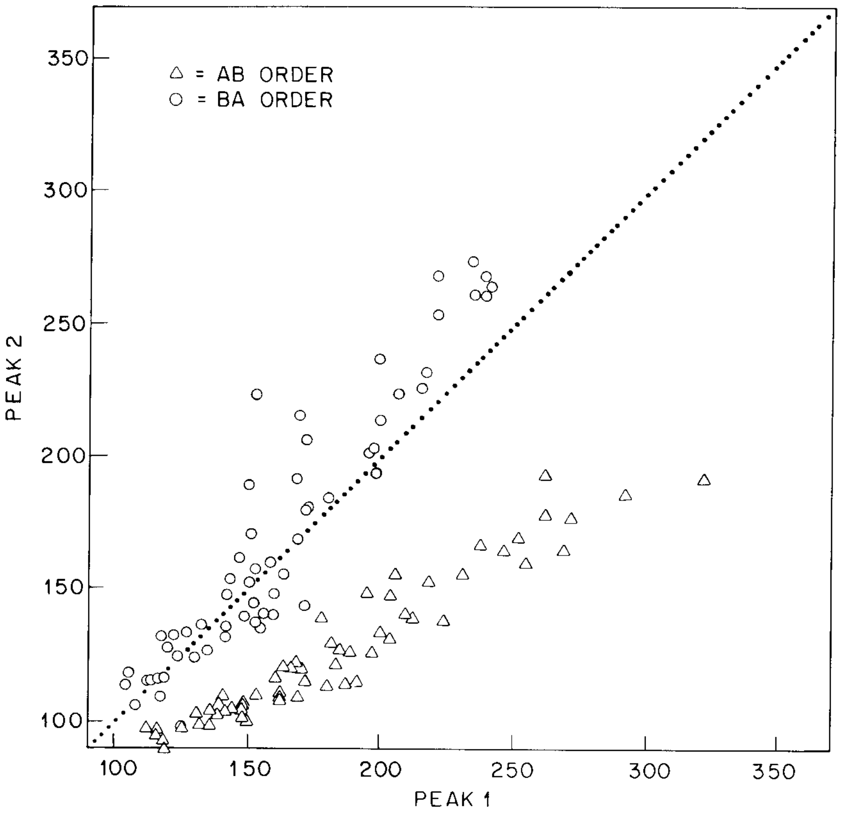


Figure 17
Answer-Background peak data for subject MYL

sequences apt to produce deaccenting, and such sequences were rearranged. As in the first experiment, the desired intonation pattern was demonstrated to subjects other than the authors, and the subject's ability to produce it naturally was checked.

The recorded utterances from both experiments were digitized, and F0 contours were computed using an LPC-based method due to Bishnu Atal. F0 measurements were made using an interactive parameter display program. In the first experiment, the measurements of primary concern were the F0 peaks of the A and B accents. Other values measured were the starting F0 in each phrase, the low point at the end of the A configuration, the valley between the peak and the terminal rise in the B configuration, and the maximum height of the terminal rise in the B configuration.

In the second experiment, the peak F0 value on each of the primary stressed syllables was measured. Only values at least two frames from a consonantal release or closure were used, to reduce the effect of transients from a [b] articulation. Since the level on one stressed syllable continues basically unchanged until just before the next stressed syllable, these measurements summarize the stepping pattern. For items before the end of the list, the F0 peak most often occurred near the end of the stressed syllable, while for the last item, it occurred near the beginning. This difference in timing seems to be related to the fact that the target on the last list item is not sustained over subsequent syllables. Instead, it is followed by a terminal fall like that in the A configuration. We also measured the lowest stable F0 in the terminal fall,[8] as well as the earliest stable F0 at the start of the utterance. For all the F0 measurements, the timepoints of the measurements were also stored.

3. Some General Characteristics of Intonational Variation

Our goal is a simple, coherent model of F0 realization that (a) interprets the categories of intonational description as we understand them, (b) incorporates the important qualitative characteristics of F0 measurements, and (c) shows good quantitative agreement as well. We will determine the model's basic outline by exploratory data analysis, including the fitting of partial or alternative models. We thus begin with some general features of the data, in order to give an intuitive appreciation for the patterns to be explained, and in order to motivate the basic concepts of our models.

Table 1
Variation in lows vs. variation in peaks, A configuration

		Peak			Low		
Subject		Mean	SD	Ratio	Mean	SD	Ratio
MYL							
	A in AB	184	49.0	.27	88	6.6	.08
	A in BA	170	50.0	.29	77	2.9	.04
	Last item in 5-list	101	18.6	.18	73	2.7	.04
	Last item in 2-list	133	35.5	.27	74	4.1	.06
JBP							
	A in AB	414	90.4	.22	158	17.8	.11
	A in BA	401	90.3	.23	142	18.4	.13
	Last item in 5-list	195	26.3	.13	148	8.7	.06
	Last item in 2-list	252	56.2	.22	148	8.9	.06
DWS							
	A in AB	260	42.7	.16	128	20.6	.16
	A in BA	257	45.1	.18	104	5.9	.06
	Last item in 5-list	128	13.2	.10	94	4.2	.04
	Last item in 2-list	170	21.3	.13	98	4.2	.04
KXG							
	A in AB	340	84.0	.25	151	22.6	.15
	A in BA	319	75.4	.24	100	16.7	.17

3.1 The Bottom of the Range Is Constant

In both experiments, increasing the pitch range had different effects on different points in the F0 contours. As figure 4 suggests, the peak values increase dramatically, while the low values remain more nearly constant. The lowest F0 values produced were quite constant over the full spread of pitch ranges. These were the low value at the end of the A configuration and the low value at the very end of the phrase in the stepping patterns.

Table 1 presents some evidence to support this claim. We examine the characteristics of phrase-final pitch accents, of the falling variety. These include the A configuration accents from the first experiment, and end-of-list accents from the second experiment. To save space, we include only data from the five-item lists and the two-item lists; the lists of length three and four behaved similarly. Data from the A configuration in utterance-medial position (the AB order) and utterance-final position (BA order) are shown separately. The data for each of four subjects are listed individually.

We have measured two points in the F0 contour of each of the falling accents under consideration: the peak value and the lowest value at the bottom of the fall. Each line of the table gives some statistics for a particular category of falling accents. For instance, the first line presents some information about the A configuration accents in the AB order sentences as spoken by subject MYL; recall that the AB order produces the pattern shown in figure 9 and arises naturally in a dialogue of the type shown in example (1).

From each set of F0 measurements (e.g., the set of peak values of A configuration accents from phrases in AB order spoken by subject MYL) we have extracted three numbers. The first is the mean, or average value, given in Hertz (cycles per second); the second is the standard deviation (SD; given in Hz), which is a measure of the amount of variation to be found in the set of numbers under consideration; and the third is the ratio obtained from dividing the standard deviation by the mean.[9] From 60 to 80 measurements were taken for each of the sets in the table.

There are several reasons for the variation observed in the data. Obviously, the most striking source of variation is the change in pitch range that our subjects obliged us by producing, but presumably smaller contributions are made by various uncertainties in production and measurement. Figure 4 leads us to expect that the variation due to pitch range manipulation will be concentrated in the values of the peaks and will have less effect on the values of the low points. Indeed, table 1 shows much lower standard deviations for the low point measurement sets than for the corresponding peak value sets. However, we cannot be entirely satisfied with this evidence—if pitch range changes involved scaling by a multiplicative constant, the effect on lower pitches would be systematically smaller than the effect on higher ones, without any implication that the bottom end of the system is "constant." If we reexpress the standard deviations as a proportion of the mean value, then any effects of simple multiplicative scaling will be removed. Again, table 1 shows that such ratios are consistently smaller for the low-point sets than for the peak sets; the only exception is the A accent in the AB order for subject DWS.

When utterances are produced in a variety of pitch ranges, then, the low-point values of falling accents show less variation than the peak values. Table 2 shows that in utterance-final position, the variation in such low-point values is relatively uncorrelated with the variation in the associated peak values.

The numbers in table 2 result from a procedure called *linear regression*, which is a way of trying to bring out the relationship between one set of

Table 2
Summary of peak-low relations for A accents

	A in AB pattern		A in BA pattern	
Subject	Slope	R squared	Slope	R squared
MYL	.04	.1	.01	.03
JBP	.13	.43	−.03	.02
DWS	.27	.32	.06	.23
KXG	.20	.54	−.06	.08

values and another. In this case, we are trying to predict the A configuration low-point values from the associated peak values; we assume that each low-point value arises from an additive constant (called the *intercept*), plus a multiplicative constant (called the *slope*) times the associated peak value. We call the relation *linear* because on a two-dimensional plot in which the peak value of a given pitch accent is plotted on the *x*-axis and the low-point value is plotted on the *y*-axis, the assumed relationship is a straight line. Of course, the actual data values do not fall exactly on a line, straight or not, but are somewhat scattered. Linear regression provides slope and intercept values that minimize the squared prediction error summed across a particular data set. We can also determine a measure of the degree of scatter in the actual data points, called *R squared*; this number gives the proportion of the variance in the predicted quantity that is accounted for by the assumed straight line.

Table 2 shows that in utterance-final position (i.e., in the BA pattern), the A accent lows do not show any consistent direction of dependence on the corresponding peaks—half of the subjects show a positive slope, and half show a negative slope. As their associated peaks rise, these final lows show little inclination to follow. Furthermore, the R squared values show that very little of the variation in these low-point values (which are not very variable to begin with) is related to changes in the associated peak values. We may conclude that the utterance-final A accent lows are essentially unaffected by pitch range changes. Furthermore, we note from table 1 that for a given speaker, the utterance-final low-point values in the first experiment are essentially the same as the utterance-final low-point values in the second experiment. This is a striking result, since the data for the two experiments were collected more than six months apart and since the intonational materials are quite different. It appears that this final low value is a relatively invariant characteristic of a speaker's voice.

The A accent lows in utterance-medial position (i.e., in the AB pattern) are rather less invariant. All four subjects show a positive slope (i.e., in all cases the lows are tending to rise as the peaks rise), and the R squared values are considerably larger than in final position, suggesting that this pattern is more consistently characteristic of the data. The results in tables 1 and 2 should be compared with those in tables 3 and 4, which list comparable statistics for the B configuration accents. Here again the variability of the low points is less than that of the associated peaks, even in proportion to their means. However, the B configuration lows have a considerably greater tendency to rise in response to the pitch range manipulation that is raising the associated peak. Both in utterance-medial and utterance-final position, the slopes and R squared values in table 4 are consistently larger than the corresponding entries in table 2.

Table 3
Variation in lows vs. variation in peaks, B configuration

		Peak			Low		
Subject		Mean	SD	Ratio	Mean	SD	Ratio
MYL							
	B in AB	128	27.4	.21	82	4.5	.05
	B in BA	162	38.5	.24	97	7.9	.08
JBP							
	B in AB	287	56.9	.20	185	28.5	.15
	B in BA	377	80.0	.21	177	25.5	.14
DWS							
	B in AB	200	36.4	.18	145	30.0	.21
	B in BA	253	38.7	.15	138	12.8	.09
KXG							
	B in AB	203	41.3	.20	141	26.9	.19
	B in BA	332	69.3	.21	174	27.4	.16

Table 4
Summary of peak-low relations for B accents

	B in AB pattern		B in BA pattern	
Subject	Slope	R squared	Slope	R squared
MYL	.10	.37	.16	.63
JBP	.42	.69	.25	.62
DWS	.70	.83	.25	.59
KXG	.59	.83	.36	.82

To sum up: lower F0 values are less affected by pitch range changes than higher ones are, and the lowest F0 values, those of utterance-final falling accents, are nearly constant for a given speaker. Similar observations have been made by Maeda (1976) and Boyce and Menn (1979) in corpus studies of F0; our results strengthen theirs, since pitch range varied much more widely in our data.

3.2 Final Peaks Are Lower

Figures 14 through 17 are plots that show the peak measurement data from the first experiment. Each figure shows the data from one subject, the *x*-axis showing the first peak values and the *y*-axis the second peak values. The utterances in AB order are plotted as △, and the utterances in BA order are plotted as ○. Thus, an AB utterance whose first accent's peak F0 value was 150 Hz and whose second accent's peak F0 value was 120 Hz would be plotted as a △ at coordinates (150, 120). In each plot, a diagonal is drawn along the line on which coordinate values are equal.

If phrasal position had no effect, we would expect the data points to be symmetrical relative to the diagonal. In other words, if the relation of the A accent peak height to the B accent peak height were independent of the order in which they occurred, then no matter how the peak heights and their relationship might be affected by pitch range changes, the resulting scatter plot should show the AB points to be the mirror image of the BA points, reflected around the diagonal.

It is clear that these plots are not symmetrical in the way just described. In general, the BA points are hardly above the diagonal at all, while the AB points are substantially below it. In other words, the second peaks are lower (relative to the first peaks) than time-order invariance would predict.

This is just the effect that theories of declination (general lowering of pitch in the course of a phrase) would predict. However, the data in figures 14 through 17 do not tell much about the nature of this declination-like effect. Since there are only two peaks per phrase, the underlying effect might actually be a lowering of the last peak, or a raising of the first one, or any revaluation of peaks throughout the phrase that leaves the last peak lower relative to the first.

Since our second experiment involved sentences containing from two to five peaks, we look to its data for clarification of this point. Recall that there were 20 different berry lists for each list length, from two-item lists to five-item lists, and that each list was read by each subject with three different pitch range instructions. One simple way to look for a phrase-position-dependent modification of F0 levels is to compare the average values in various positions in lists of different lengths.

Figures 18 through 20 show the data for three subjects, averaged by pitch range instruction and list length. The *x*-axis shows position in the list, counting from the beginning, and the *y*-axis shows average F0 (in Hz).

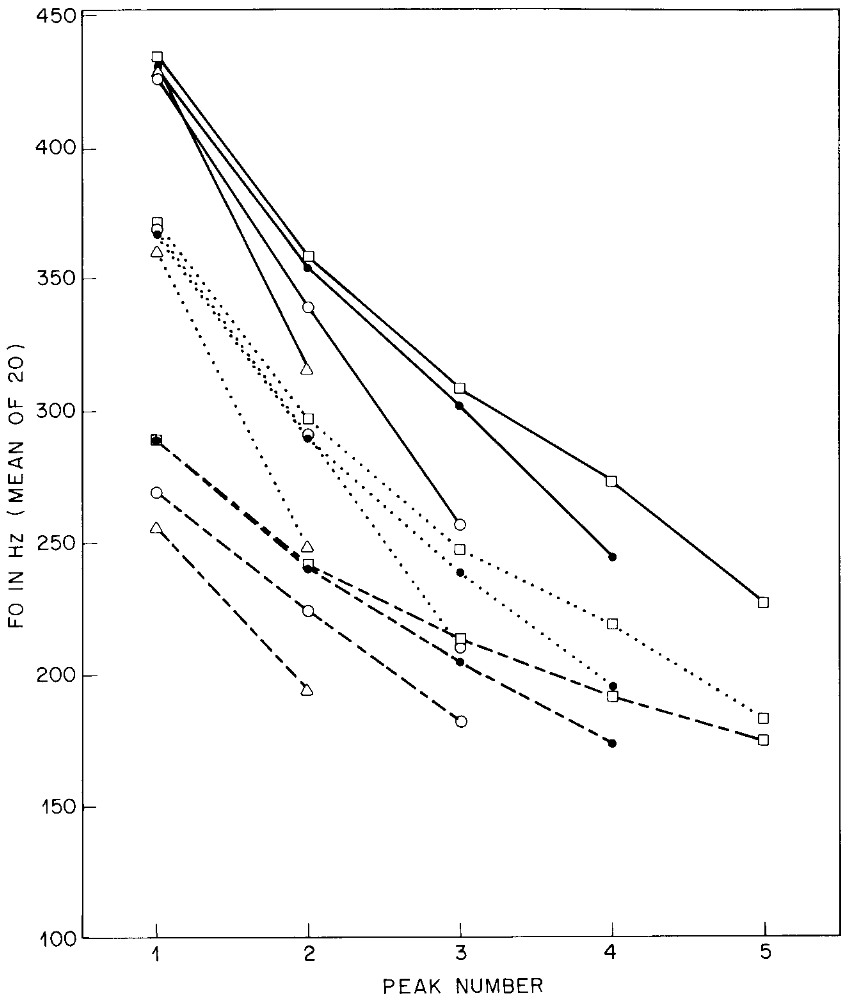


Figure 18
Downstep data (3 pitch ranges, 4 lengths) for subject JBP

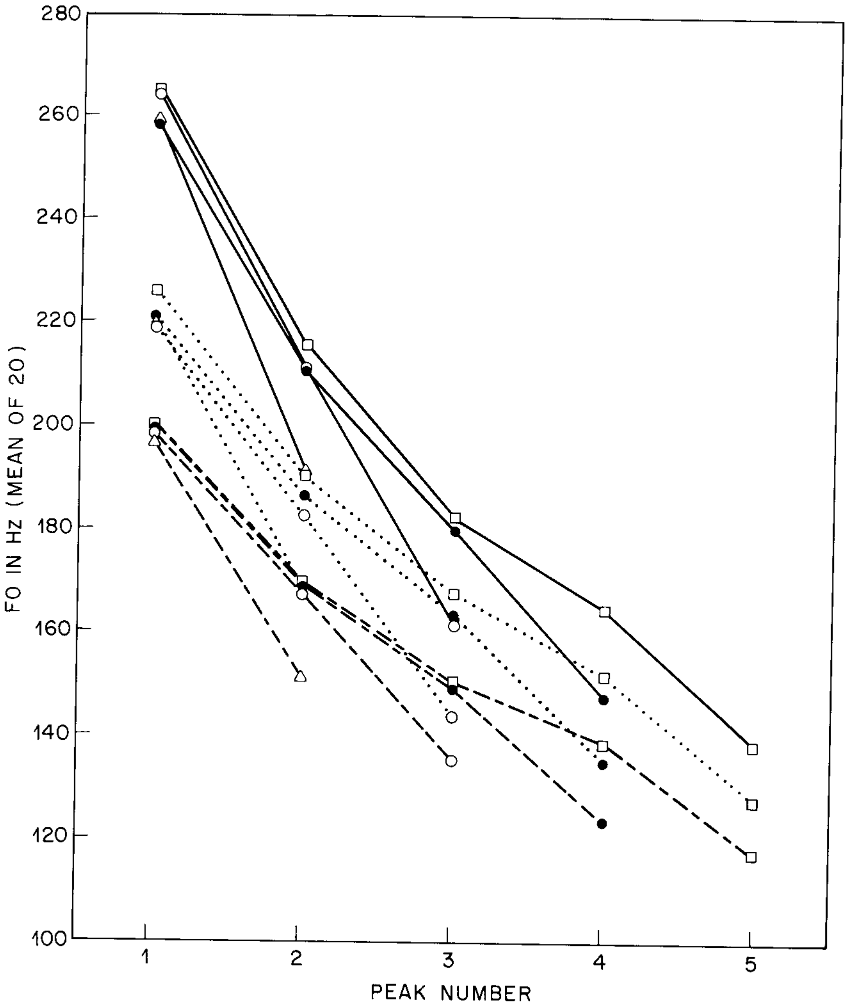


Figure 19
Downstep data (3 pitch ranges, 4 lengths) for subject DWS

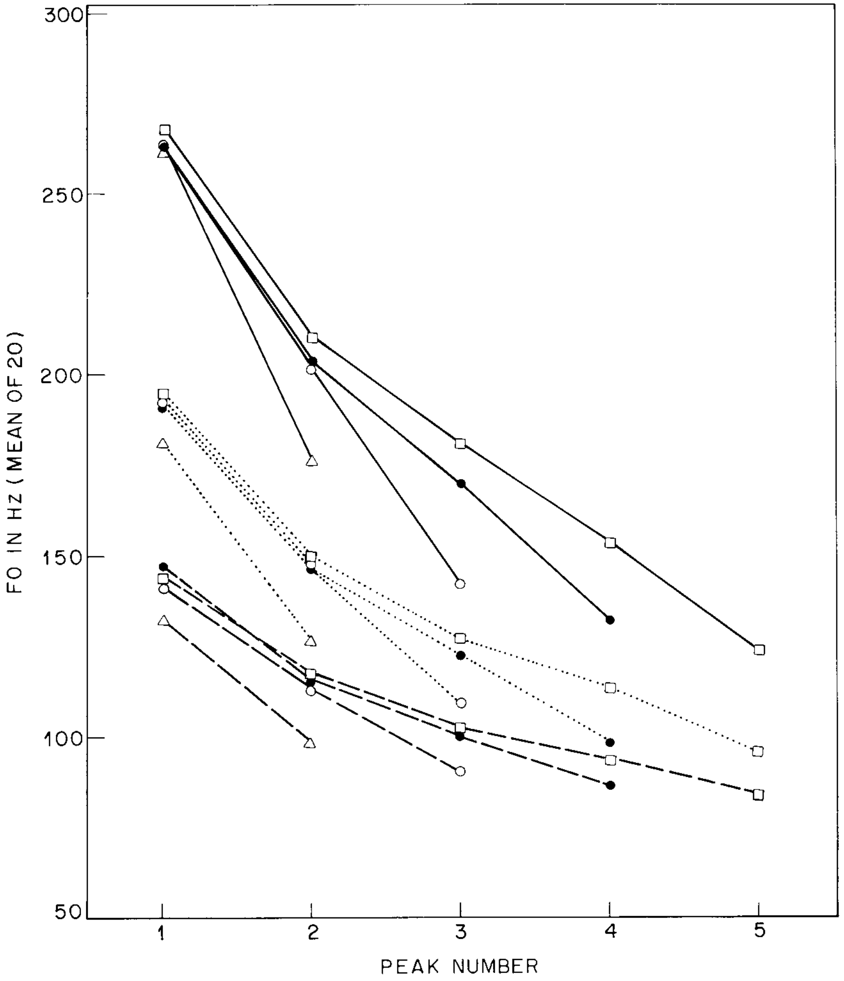


Figure 20
Downstep data (3 pitch ranges, 4 lengths) for subject MYL

Solid lines connect the measurements for pitch range instruction 3, dotted lines connect those for pitch range 2, and dashed lines connect those for pitch range 1. The plotting character □ is used for five-word lists, ● for four-word lists, ○ for three-word lists, and △ for two-word lists.

All the lines have a clear downward trend, which is as it should be, since the subjects were supposed to produce downstepping contours; it is also clear that the subjects have produced downstepping patterns in three pitch ranges, as instructed. There is some tendency for the shorter lists in a given pitch range to begin lower. Against the background of these general patterns, the most striking effect of phrasal position is to be found in the list-final positions: the list-final measurements are generally lower than might be expected.

Two different aspects of this final lowering are visible in figures 18 through 20; the first depends on extrapolating the trend of the nonfinal measurements, and the second depends on comparing the final measurements with nonfinal measurements in the same serial position in longer lists. In order to display these effects more clearly, figure 21 shows just the pitch range 3 data for subject DWS. If we were to see just the first four points of the five-item list in figure 21 and were to try to predict the fifth one from them, where would we expect it to fall? The answer to this obviously depends on what kind of pattern we think is to be extrapolated. If we look at these four points, and in general at the nonfinal points in all the plots in figures 18 through 21, it seems clear that they do not form straight-line patterns. Instead, they seem to fall on curves that are concave upward. A reasonable guess about the nature of these curves would be that they are exponentially decaying series—that is, sequences in which each number is a constant fraction of the previous one:

(3)
Exponential decay

$$X_{i+1} = s \cdot X_i$$

Where s is a constant less than 1, equation (3) describes the successor relation in such a series. Such sequences fall to nearly zero quite quickly—they asymptote to zero—unless a nonzero asymptote is provided. This is easily done as shown by the following equation for asymptote r:

(4)
Exponential decay to a nonzero asymptote

$$X_{i+1} - r = s \cdot (X_i - r)$$

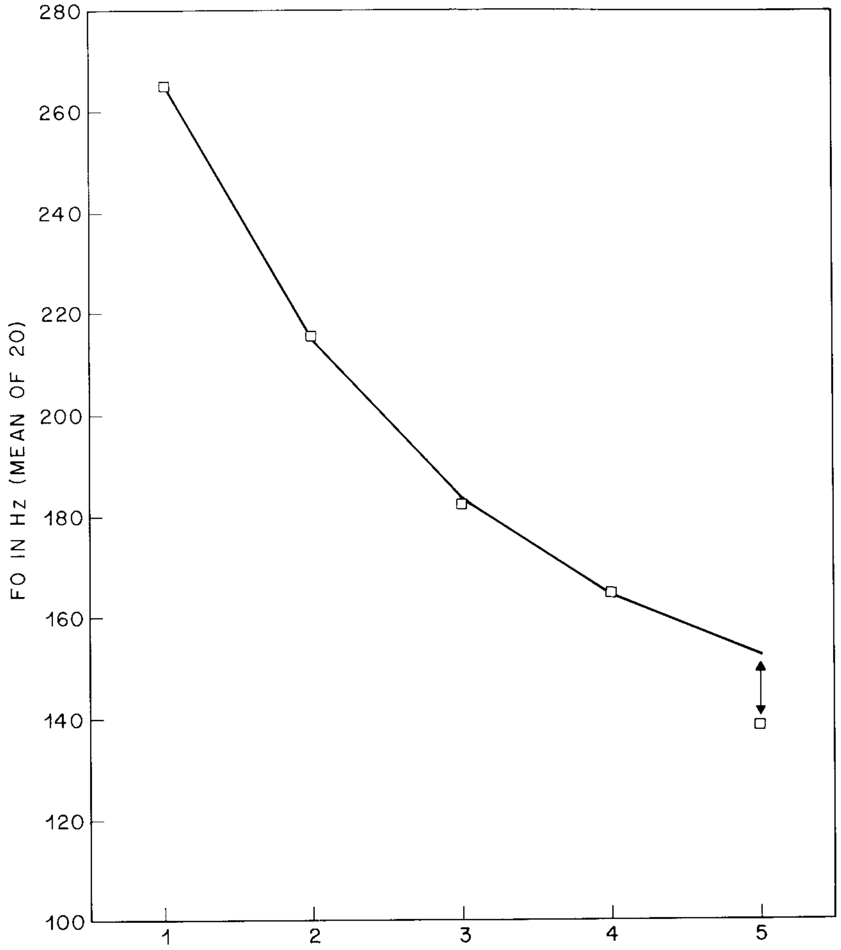


Figure 21
Pitch range 3, length 5 data for subject DWS. The solid line is a decaying exponential fit to the first four points of the five-item list. The arrow indicates how far the fifth data point falls below the value predicted by this exponential.

Table 5
Errors in simple downstep model: Predicted – Actual (in Hz)

Subject	Pitch range	Item 1	Item 2	Item 3	Item 4	Item 5
MYL						
	PR3	0.6	0.4	−4.1	3.1	21.2
	PR2	−1.0	2.0	0.1	−1.1	7.8
	PR1	−0.3	0.5	0.1	−0.2	4.3
JBP						
	PR3	0.7	−1.2	−1.1	1.8	27.3
	PR2	−0.2	−0.4	1.3	0.5	15.4
	PR1	0.1	0.2	−1.4	1.3	5.5
DWS						
	PR3	0.1	−0.9	1.4	−0.4	13.8
	PR2	0.3	−0.5	−0.7	1.0	16.3
	PR1	0.0	−0.2	0.2	0.1	14.2

In figure 21 the solid line represents a decaying exponential fit to the first four points of the five-item list. It fits the first four points rather well, but the fifth point, the last one, falls quite far below it (its shortfall is indicated by a double-headed arrow). According to this method of extrapolating the trend of the nonfinal points, the last measurement in all of the four-item and five-item lists is lower than expected. (We cannot try this particular test with the three-item lists, since the two nonfinal elements do not provide enough information to estimate the necessary parameters.) Table 5 summarizes the results of this enterprise for the five-item lists. We have fit each subject separately; for each subject, we have allowed each pitch range to choose its own asymptote, but have required the same downstep constant for all pitch ranges.

We have thus created a crude model of downstep: it assumes that downstep is exponential decay, once we have subtracted the asymptote or "reference level"; and it assumes that each of our pitch range instructions is associated with a specific "reference level." By numerical methods, we produce (for each subject) those values of the downstep constant and the three reference levels that will make the model fit best to 12 numbers—i.e., the four nonfinal average F0 measurements in the three pitch ranges.

To obtain the numbers in table 5, we subtract the average measured F0 value, in each list position, from the values predicted by the model, showing the errors in the model's predictions. The errors in the first four columns are fairly small, while the errors in the fifth column are larger; this is not surprising, since the model parameters were selected so as to fit the first four columns of numbers, ignoring the last column completely. The important point about the numbers in the last column is that they are positive, indicating that the final measurements fall reliably below the trend extrapolated from the nonfinal measurements.

Table 6
Differences (in Hz) between averaged lists of different lengths, by position

Comparison	Position 1	2	3	4
5–4	3	4	5	18
5 3	6	8	27	
5–2	10	31		

A second way to see how the data in figures 18 through 21 point to a special lowering in final position, and one that does not depend on any assumptions about the nature of the downstep function, is to compare the final measurements in the four-, three-, and two-item lists with the nonfinal measurements in the corresponding position in longer lists. In figures 18 through 20, compare the final and nonfinal values in the second, third, and fourth serial positions. The final vs. nonfinal comparisons show consistently and substantially lower final values, an effect that is larger and more consistent than the lowering of shorter lists in nonfinal positions.

A crude numerical summary of this observation can be produced by subtracting averaged four-, three-, and two-item lists, position by position, from averaged five-item lists. The results are shown in table 6 (here we have averaged lists of each length across subjects and pitch ranges). The previously noted tendency for shorter lists to start a little lower is visible in the increasing values (3, 6, 10) in the column corresponding to the first position. A very small amount of "downdrift" is hinted at by the sequences (3, 4, 5) and (6, 8) for the differences in nonfinal positions in the 5–4 and 5–3 comparisons; however, these small increases could easily be due to chance variation in the data. The largest and most striking effect in the table is the evidence for final lowering; in each row, the last number is much the largest.

3.3 "Speaking Up" Raises the "Reference Level"

How should we model the consequences of "speaking up"? With respect to the downstep data, given the general idea that downstep is decay to an asymptote, we have three obvious options: "speaking up" could (a) increase the reference level, (b) increase the starting value, or (c) change the decay constant. Various combinations of these choices (and of course there are many other possible approaches) will result in different models. In deciding which models to pursue, we will consider both the patterns in the data, insofar as we can be sure what they are, and some general ideas about how such models should work.

In this case, we would like to separate the modeling of downstep cleanly from the modeling of pitch range changes. Therefore, if the data permit, we would like to maintain the downstep constant independent of pitch range changes, since this parameter is peculiar to the modeling of downstep. If the reference level were also an intrinsic aspect of the downstep phenomenon, we would also be inclined to insulate it from pitch range manipulation. However, it is straightforward, if we wish, to treat the reference level as an aspect of F0 scaling in general, thus making it a reasonable parameter for pitch range manipulation to modify. In deciding how to proceed, we should look to the data for guidance.

In computing the numbers in table 5, our optimization procedure chose a different asymptote value for each pitch range. These asympotoes, or reference levels, were simply chosen to produce the best fit to the data—no assumptions about reference levels rising with pitch range instruction were made. If the resulting values were relatively constant, showing no clear tendency to rise with pitch range, we would be free to make the reference level part of the downstep rule, using (for instance) the starting value as the principal correlate of pitch range (of course, a constant asymptote might still result from more general characteristics of the F0 system). However, visual inspection of figures 18 through 20 suggests that the ultimate limit of downstep is probably higher in higher pitch ranges, and the results shown in table 7 provide quantitative support for this view. The increases in estimated reference level are sufficiently large, and sufficiently consistent, for us to adopt the idea that if our model is to have a reference level parameter, it should increase with pitch range.

Table 7
Asymptotes fitted to nonfinal measurements in 5-item lists

	Pitch range 1	Pitch range 2	Pitch range 3	Downstep constant
MYL	81	91	128	.59
JBP	158	163	218	.64
DWS	120	130	133	.62

Table 8
Differences between initial peak values and estimated reference levels in Second Experiment

	Pitch range 1	Pitch range 2	Pitch range 3
MYL	63	104	140
JBP	130	208	216
DWS	80	86	132

3.4 "Speaking Up" Raises the Initial Pitch with Respect to the Reference Level

Without question, increasing the pitch range increases the F0 value of an initial peak or stepping accent. Once reference levels are taken out, do the initial peak values still increase? The data in table 8, derived from the same analysis as the data in tables 5 and 7, suggest that they do.

4. Modeling F0 Implementation

In the previous sections, we have sketched a number of characteristics of the F0 patterns in two data sets. Now we will try to show that the various particulars can be combined into a coherent model.

4.1 A Simple Model

The first model we will try involves five basic principles, with seven parameters that are constant for all utterances of a given speaker and one parameter (representing pitch range) chosen for each phrase. These parameters and rules are exact hypotheses about the nature and interactions of the general characteristics noted above.

Our approach will hypothesize an F0 *transform*, which translates measured F0 values into a new set of values that are assumed to behave in a simpler way. If the approach is a correct one, the domain of transformed F0 values should bring us closer to the underlying phonetic control parameters for intonation. The answer-background relation and the downstep relation are each taken to be a constant ratio in transformed F0 values, with a subsequent lowering of F0 targets in utterance-final position. We use the symbol k for the answer-background constant and the symbol s for the downstep constant. The transformed value of a given F0 target P is taken to depend on pitch range; we assume that a reference level r is set for each phrase, the transformed value of an F0 target P being its distance above r. Since we observed (in section 3.1) that utterance-final low points have a fixed value that does not increase with pitch range, it seems plausible to consider this low F0 value as the bottom of the entire system, constraining r to always remain above it. We accomplish this aim by requiring r to remain at least d Hz above this low value, which we symbolize as b. The effect of phrasal position is assumed to be limited to lowering in final position, and the lowering effect is assumed to be a constant fraction of the distance of P above r. We use the symbol l for the final lowering constant.

(5)
Model 1

a. General F0 transform
$T(P) = P - r$
P and r in Hz

b. Downstep
$T(P_i) = s \cdot T(P_{i+1})$
where P_i is the F0 target in Hz of a step accent in position i, downstepped with respect to the previous accent target P_{i-1}

c. Answer-background relation
$T(P_A) = k \cdot T(P_B)$
where P_A is the F0 target in Hz of the A accent, and P_B is the target of the B accent

d. Relation of r to initial accent target
$r = f \cdot (P_0 - b)^e + d + b$
where P_0 is the target in Hz of the first pitch accent, and d, e, f, and b are constants

e. Final Lowering
$P \rightarrow r + l \cdot (P - r) / \_\_\_\_ \$$
where $l < 1$

We have chosen the somewhat mixed formalism used above with no aim in view, for the moment, but to express our general observations with sufficient precision to constitute a testable model. The equations in (5a–d) express a set of constraints to be satisfied. (5e) is expressed as a pseudo phonological rule, since the values of P on either side of the arrow are different, the sense being simply that F0 targets are lowered in phrase-final position. This choice of symbolism has no aim beyond clarity; note, for instance, that the identical model could be reexpressed by deleting rule (5e) and modifying the general transform in (5a) to read as follows:

(6)
Modified transform for model 1
$T(P) = (1/l) \cdot (P - r)$
where $l < 1$ in final position, $l = 1$ otherwise

We chose the form in (5) in order to separate the claim of final lowering from the claim that F0 relations are constant ratios once r has been subtracted.

The equation in (5d) is a frankly empirical approximation to the interrelationship of the reference level r and the initial accent target P_0. Equation (5a) assumes that both P and r are known, while equations (5b) and (5c) express only ratios of adjacent accent values. Therefore, without an equation like (5d), the model would need two numbers specific to each utterance: the reference value r and one of the accent targets, say the first one. Equation (5d) suggests that these two numbers co-vary in a predictable way, together expressing the speaker's response to the pitch range instruction, so that only one parameter need be fixed in modeling an individual utterance. We put the exponent e in equation (5d) because preliminary modeling like that represented in tables 7 and 8 suggests that the relation is nonlinear.

4.2 Fitting a Model

According to the model we have just described, if the constants k, s, l, b, d, e, and f are correctly set, then knowledge of any one of the peak F0 measurements in a given utterance allows us to predict all of the other peak F0 measurements in that utterance. The difference between these predicted values and the actual data gives us a measure of the model's fit, and we can use hill-climbing techniques to find the parameter values that give minimum error.

For each type and length of utterance, given that values have been assigned to the model's constants, each value of the pitch range parameter r (or equivalently P_0) predicts a set of F0 targets. For utterances of N accents, this function from r-values to sets of accent targets describes a curve in N-space. The F0 measurements for a particular N-accent utterance specify a point in N-space; the corresponding set of predicted F0 values can then be found at the place on the model's curve nearest to that point, and the error measure for that utterance could be the euclidean distance from the point to the curve, or the mean squared difference between the predicted values and the measured ones, or whatever. In the exposition that follows, we will use the average absolute value of the

Table 9
Mean absolute error, in Hz, model 1

	Downstep data	AB data
MYL	2.2	4.9
DWS	1.5	5.4
JBP	4.8	7.1
KXG		7.8

differences between model values and data values as our error measure. This measure is easier to understand than euclidean distance in N-space, and it does not emphasize outliers in the way that squared error does.

We fit each subject separately, but both experiments at once. The constant k is relevant only to the Answer-Background experiment, and s only to the Downstep experiment, but the five other constants help predict both data sets. For the Answer-Background experiment, we fit our models to the raw peak F0 measurements, but for the Downstep experiment, we first averaged each subject's peak F0 measurements by length and pitch range instruction, since the variation of berry names across positions allowed this procedure to remove segmental effects on the measurements. Each subject provided 42 data points in the Downstep experiment, each data point being the average of 20 observations, and 160 data points in the Answer-Background experiment. In order to combine the error measures for a given set of parameter values tested against the two data sets simultaneously, the average absolute error was computed for each data set independently, and the two averages were summed.

Table 9 summarizes the fit of model 1 to the data, and figures 22 through 28 demonstrate it graphically. Figures 22 through 25 illustrate the fit of model 1 to the data from the first experiment, by subject. As in figures 14 through 17, △ represents the AB data points and ○ represents the BA data points. The solid lines are the predictions of the model for the two cases, derived by fitting the model to the data from both experiments at once. Figures 26 through 28 illustrate the fit of model 1 to the data from the second experiment, by subject. As in figures 18 through 20, □ is used for data points from five-word lists, ● for four-word lists, ○ for three-word lists, and △ for two-word lists. Solid lines show the model predictions for pitch range instruction 3, dotted lines show those for pitch range instruction 2, and dashed lines show those for pitch range instruction 1. Again, these predictions are derived by fitting the model to the data from both experiments at once. Table 10 gives the corresponding values for the seven constant parameters. Given the amount of scatter in the data, the error

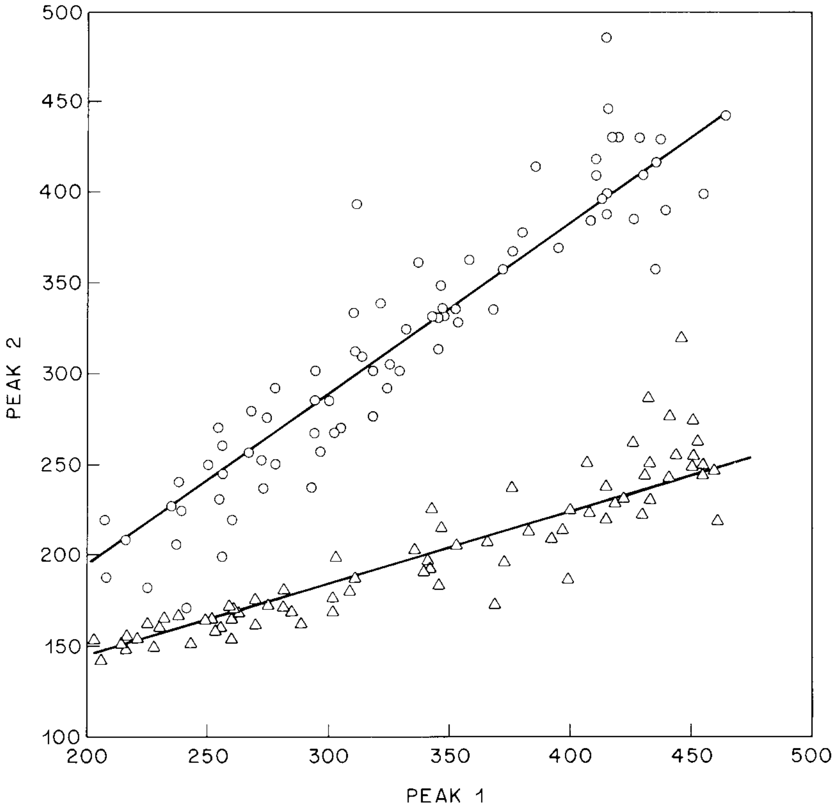


Figure 22
Model 1 fit for subject KXG

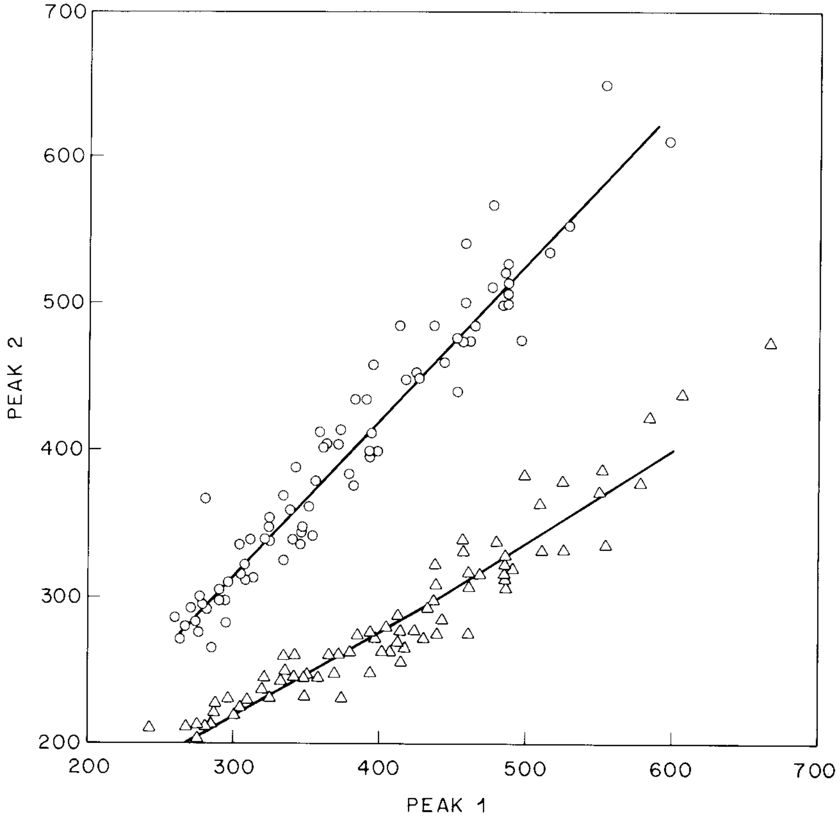


Figure 23
Model 1 fit for subject JBP

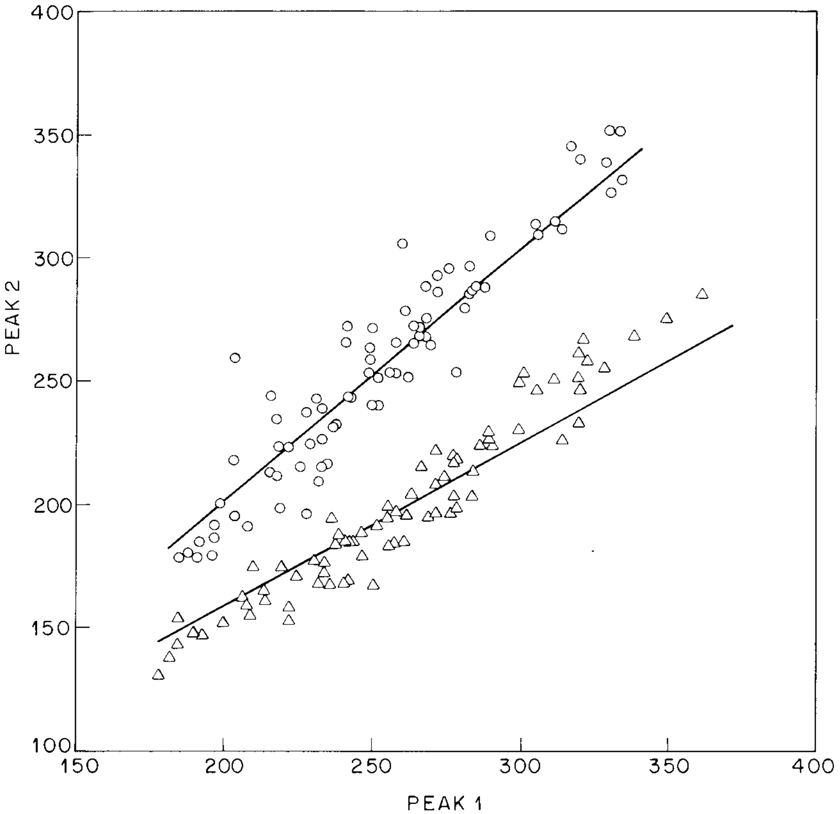


Figure 24
Model 1 fit for subject DWS

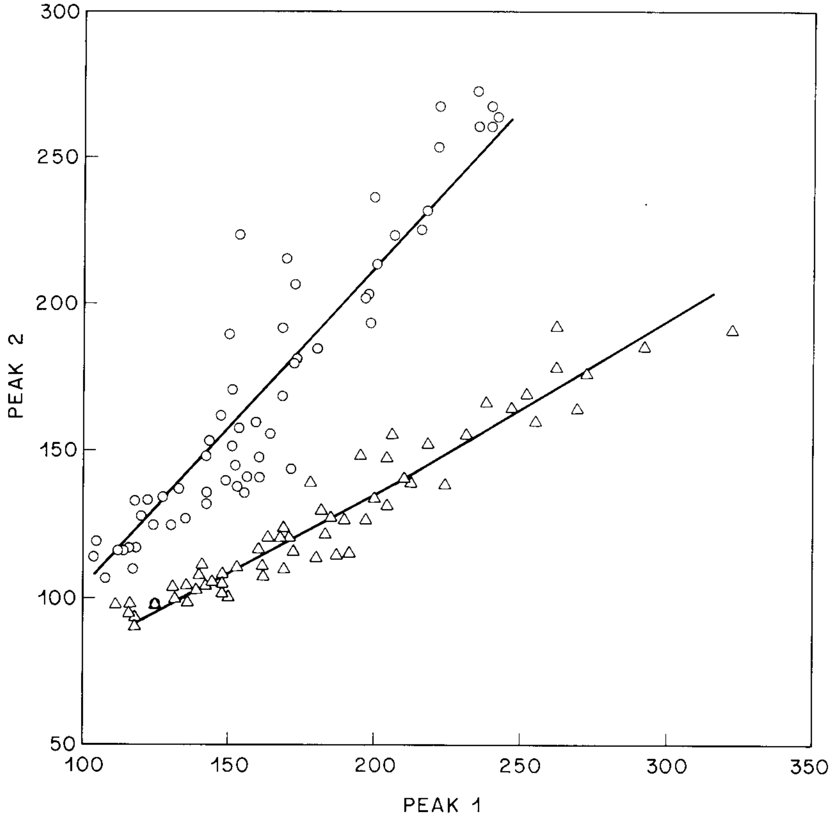


Figure 25
Model 1 fit for subject MYL

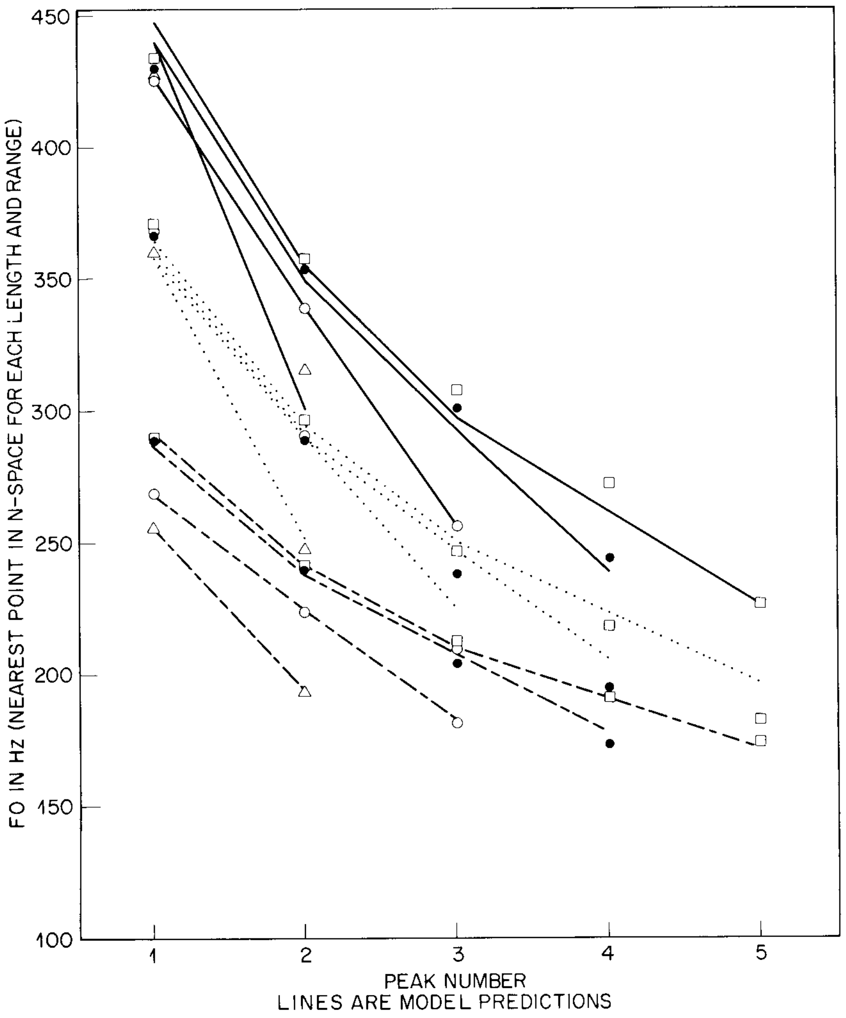


Figure 26
Model 1 fit for subject JBP

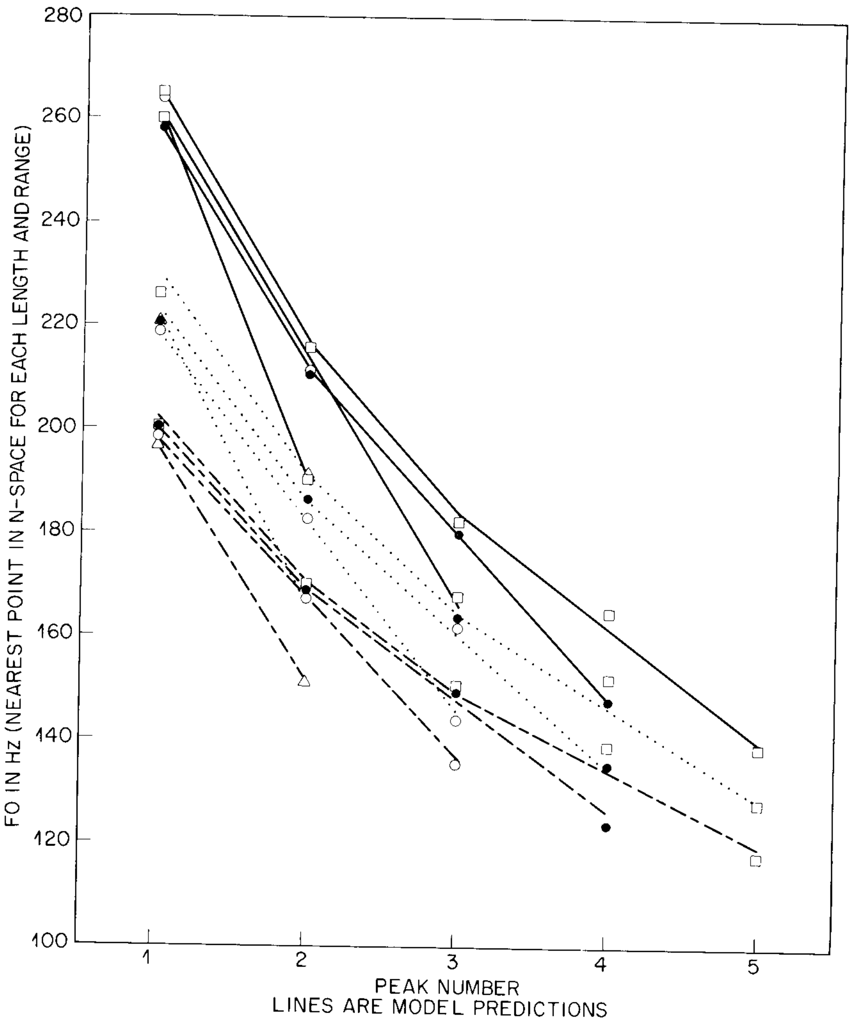


Figure 27
Model 1 fit for subject DWS

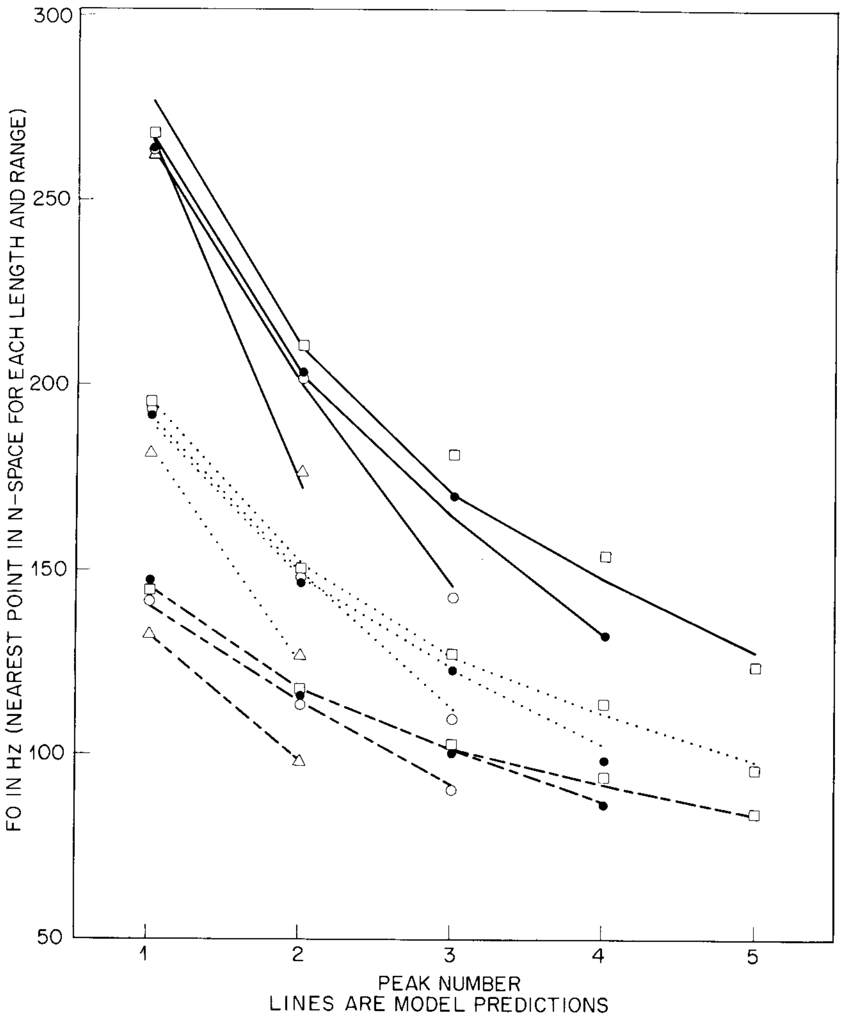


Figure 28
Model 1 fit for subject MYL

Table 10
Parameter values for fit of model 1

	s	k	l	f	e	b	d
MYL	.59	1.66	.68	.0059	1.67	64.3	4.9
DWS	.68	1.33	.77	.0049	1.63	81.3	9.7
JBP	.62	1.63	.68	.0049	1.64	111.9	21.8
KXG		1.59	.59	.0049	1.33	90.3	18.9

measures in table 9 are encouragingly small. We may conclude that our initial observations, which gave rise to the model, describe the data fairly well, and that no large surprises emerge from the interaction of the various effects we have posited. Of course, this does not mean that the model is correct. There may well be small effects that are not covered; there are many rather different models that make nearly equivalent predictions about our data sets; and there may be new phenomena, or different ways of measuring the phenomena we have examined, that require a different approach.

Just as importantly, the meaning of such models needs to be clarified. Which aspects of such a model are approximations to the universal physiology of F0 control, and which are facts about English or about the speech habits of American intellectuals? We have interpreted phenomena such as downstep at the level of phonetic implementation, since the objects of interpretation are phonological representations, but the interpretation process is modeled by arithmetic over continuously valued features. If the interpretation process is not universally given, then we owe an explanation of what its free parameters are. For whatever aspects of it are indeed universal, a detailed psychological and biological account is in order. In any case, the boundary between phonological description and phonetic implementation needs very careful scrutiny, since most instances of allophonic variation can easily be recast as context-dependent phonetic implementation.

The fundamental questions are only coherent when taken together: what are the basic entities, how do they combine, and what are the observable consequences of particular combinations of such entities? We find it unfortunate that (for historical reasons) the first two questions have been the principal domain of phonology, while the last one has been taken to be the (largely independent) concern of phonetics. One of our aims has been to demonstrate the value of a unified approach.

5. Some Alternative Models

In deciding on the details of model 1, we made a number of choices about the nature of the phenomena under study. Some of these choices were dictated by the patterns in our data: for instance, the data suggested that we view downstep as decay to a nonzero asymptote r (rather than linear stepping in Hz or decay to zero); the data argued that r should increase with pitch range; and the data required an extra lowering in last position in the downstepping contours. Other choices were motivated by a desire for generalization: for instance, we used the same $P - r$ F0 transform for the two data sets, even though the Answer-Background experiment provides no direct evidence for the existence of a reference level that increases with pitch range; and we equated the time-order asymmetry in the Answer-Background experiment with the final lowering visible in the Downstep data. A third set of choices was essentially arbitrary: for instance, we designed the equation relating r and P_0 just so as to make the rest of the model work, since the patterns suggested by the Downstep experiment (in tables 7 and 8) at best only specify the function at three points, and since the Answer-Background data offer little direct guidance. In many cases, we could have made quite different choices and still ended up with a model that fit about as well, but some alternatives would have been disastrous. In this section, we will consider a number of alternative approaches, good, bad, and indifferent.

5.1 Evaluating a Model

There are several different ways to evaluate a model. The crudest evaluation is the error measure: how well does it fit? This measure has no meaning without some standard of comparison: how well should it fit? How does the error compare with the error implied by the apparent variability of the data? How well do other models fit? A more sensitive evaluation is provided by looking for patterns in the prediction errors, or "residuals." Systematic errors suggest problems in the model.

Another important set of questions concerns the number and nature of the model's parameters. With a sufficient number of arbitrary parameters, any data can be fit arbitrarily well. The smallest reasonable number of parameters should be used, and the true contribution of each one to the fit should be honestly assessed.

The last, best question about a model is how well it copes with new material, in domains far from those that spawned it. We will list some areas where experiment is likely to clarify and strengthen our assumptions, or else force their retirement.

5.1.1 Evaluation of Model 1 It should be clear from the average error values in table 9 and the plots in figures 22 through 28 that the model fits quite well.

There is no obvious pattern in the residuals that is consistent across subjects. In the Downstep data, JBP and MYL show a less steep fall, for the highest pitch range in four- and five-item lists, than the model predicts, but DWS does not show this effect. For the middle pitch range, JBP consistently shows a steeper fall than the model predicts, but MYL and DWS do not show such an effect. These deviations from the model's predictions in the Downstep data are caused entirely by a problem with the function relating r and P_0. For JBP, for instance, table 7 shows that the length-five data indicate asympototes at (158, 163, 218) Hz; the parameters derived from fitting model 1 instead put these three asymptotes at (157, 178, 197) Hz. The pattern of error in modeling JBP's Downstep data thus results from the fact that the function relating r to P_0 does not become sufficiently nonlinear, presumably because the consequences in fitting the Answer-Background data would be undesirable.

If the function relating r to P_0 were linear, then the functions relating the two peak F0 values in the AB experiment would also be linear. The nonlinearity of equation (5d) causes these functions to be curved. The exponent e in this equation is the main determinant of the direction of curvature. Roughly speaking, $e > 1$ predicts that the data from the Answer-Background experiment will show a "wishbone" configuration in plots like those of figures 14 through 17, while $e < 1$ predicts a "trumpet" shape. The data from the two experiments agree in setting e greater than 1. However, given all the other details of model 1, setting the exponent in equation (5d) as high as JBP's Downstep data would suggest makes the model for the AB/BA data have excessive curvature in the higher pitch range area.

Still, in the Answer-Background data, an appropriate curvature or tilting of the model's predictions would make for a slightly better fit. There is a tendency for the low pitch range BA peak 2 predictions to be a little too high, for instance.

The residual effects just noted are fairly small, and they have several possible interpretations. The two experiments are different in many ways, and it is possible that as a result they should for instance have different relations between r and P_0 or different final lowering functions. According to our theory, r and P_0 should be freely variable, in the general case—they co-varied in our experiments only because our subjects chose (unconsciously) to adopt a certain strategy for interpreting our command to vary pitch range. Thus, the different circumstances of the two experiments might have led to a difference in strategy for these two subjects.

It is also possible that these patterns in the residuals point to some flaw in the model or the theory that underlies it. For instance, it may be that the Answer-Background relation should be modeled as a change in the reference line r, rather than as a change in the height of the F0 target above it; the details of the result would depend very much on other choices, but obviously this move could greatly alter the effects of the P_0-to-r relation on the model's predictions for the Answer-Background data.

Without any doubt, model 1 has too many parameters, in the sense that the available data underdetermine their values. We have devoted four parameters to the function relating P_0 and r, although the data constrain this function relatively little, and the two data sets do not, it seems, entirely agree about what its shape should be. As stated earlier, we adopted the form (5d) in the belief that the reference line r would have to stay somewhat above the baseline b, that its relation to P_0 might be nonlinear, and that perhaps only the portion of P_0 above b should be relevant in determining r. Except for b, which is identified with the speaker's invariant final low F0 value, none of the parameters in this equation have any clear interpretation. The parameter d is the translation of "somewhat," while e and f are just a way of getting a curved function with a minimum number of additional parameters.

Unfortunately, these four parameters can trade off against each other in various ways to give very similar functions. In the case of the Downstep experiment, even though 12 asymptotes are in fact required to fit the data, these group into three classes determined by pitch range instruction. In the case of the Answer-Background experiment, changes in the value of the constant k can somewhat compensate for the effects of P_0-to-r relation.

The parameter b is supposed to represent utterance-final low values; a comparison of tables 1 and 10 shows that b is consistently too low, but that the sum of b and d is much closer. The parameters k, s, and l are interpretable, in the sense that our model asserts certain relations among F0 values to be constant ratios in an appropriately transformed space. However, we have no a priori idea of what these ratios should be, so that the values that emerge from the modeling cannot be checked against independent evidence.

5.2 Some Workable Alternatives

It should be obvious from the preceding discussion that model 1 could be changed in various ways without particularly affecting its goodness of fit. Some of these changes are minor ones. For instance, the function relating r to P_0 could be simplified to involve only three parameters:

(7)
Model 1A

Substitute

$r = f \cdot (P_0)^e + d$

for equation (5d) in model 1.

A further simplification reduces the number of parameters to two:

(8)
Model 1B

Substitute

$r = f \cdot P_0 + d$

for equation (5d) in model 1.

The results of fitting these modified models are shown in table 11. These changes have relatively little effect on the Answer-Background data fit, but the fit to the Downstep data becomes consistently worse, due to an exaggeration of the error pattern noted in the previous section.

Another set of changes to model 1 involves equation (5e), which implements Final Lowering.

(9)
Model 1C

Substitute

$P \rightarrow l \cdot P / \_\_\_\$$

for rule (5e) in model 1.

As table 12 shows, the error measures for model 1C are similar to those for model 1. Examination of the residuals does not show any crucial advantage for one version of Final Lowering over the other.

Some alternative models that fit fairly well differ more profoundly. We previously alluded to the possible treatment of the Answer-Background relation as a change in r; for reasons of space, we will not pursue this here. Another class of models provides for the treatment of "low" tones, which do not arise in the data we have been modeling. In the F0 contours we have been working with, increases in emphasis make F0 targets go up. However, there are some cases in which the opposite happens, and increased emphasis causes lowering of F0. For this reason, and for other reasons detailed in Pierrehumbert (1980), it is reasonable to postulate two tonal categories in English, represented as H and L (for "high" and "low"). According to the transform given in (5a), the zero value for H tones is at r, with increasing values going up from there. What should we expect to happen for L tones?

Table 11
Mean absolute errors for models 1A and 1B

	Model 1A		Model 1B	
	DS data	AB data	DS data	AB data
MYL	2.2	4.9	4.0	4.8
DWS	1.8	5.5	1.8	5.8
JBP	5.3	7.0	5.9	7.0
KXG		7.8		7.8

Table 12
Error measures for model 1C

	DS data	AB data
MYL	2.2	5.1
DWS	1.5	5.1
JBP	5.6	8.0
KXG		7.9

We know that such tones are constrained to remain above the baseline b; it is reasonable, if not strictly necessary, to assume that an L tone will usually be lower than an H tone in comparable circumstances and we have just observed that L tones seem to scale downward under emphasis. All of this suggests a transform in which H tones scale upward from r, and L tones scale downward from r, with b forming a floor below which the L tones cannot be pushed. For instance, the following transform gives such a result:

(10)
Model 2

Substitute

$T(P) = \log((P - b) / (r - b))$

for equation (5a) in model 1.

With this equation, the transformed values of H tones range from 0 to positive infinity as their phonetic realizations range from *r* to positive infinity. The transformed values of L tones range from 0 to negative infinity as their phonetic realizations range from *r* down to *b*. The relation of transformed to untransformed values for two different choices of *r* is illustrated in figure 29.

The strong nonlinearity of the hypothesized relation between transformed and untransformed values is what makes it possible for L and H tones to be treated symmetrically in the transformed domain, while behaving quite unsymmetrically in the F0 domain. Many other nonlinear transforms could be found that would have the same property. If any transform in this class fits data for both L and H tones, then the symmetry in the transformed domain could be exploited to make the rules for phonetic realization of tones more general. Specifically, prominence values might be implemented as ratios of absolute values in the transformed domain, regardless of the types of the tones involved. The sign of the transformed values would then be determined by the tone type. This scheme for implementing prominence relations would generalize our equation (5c) to cover strings of free L accents and mixed configurations of H and free L accents.

Our experiments did not address the behavior of tones taking on negative values under transform (10). However, our peak and step data can be used to investigate the consequences of the scaling on the positive side. The results of fitting model 2 to the AB/BA and berry list data are shown in table 13.

5.3 Some Models That Don't Fit

The preceding section may have given the impression that any model at all will fit our data. This is by no means true. There are many alternative ways to express the basic phenomena we have found, but a model that neglects to express some relevant regularity, or a model expecting nonexistent patterns, will fail badly.

In particular, we have verified the failure of models that leave out final lowering, models that do not provide a parameter like *r* to represent the increasing asymptote of downstepping patterns with increasing pitch range, models that do not provide a prepausal low value that is invariant under changes in pitch range and length, and models that predict a rate of declination that varies directly with phrase length. Such models show poorer fits; more importantly, they show a pattern in the residuals that reflects the omitted or added assumption.

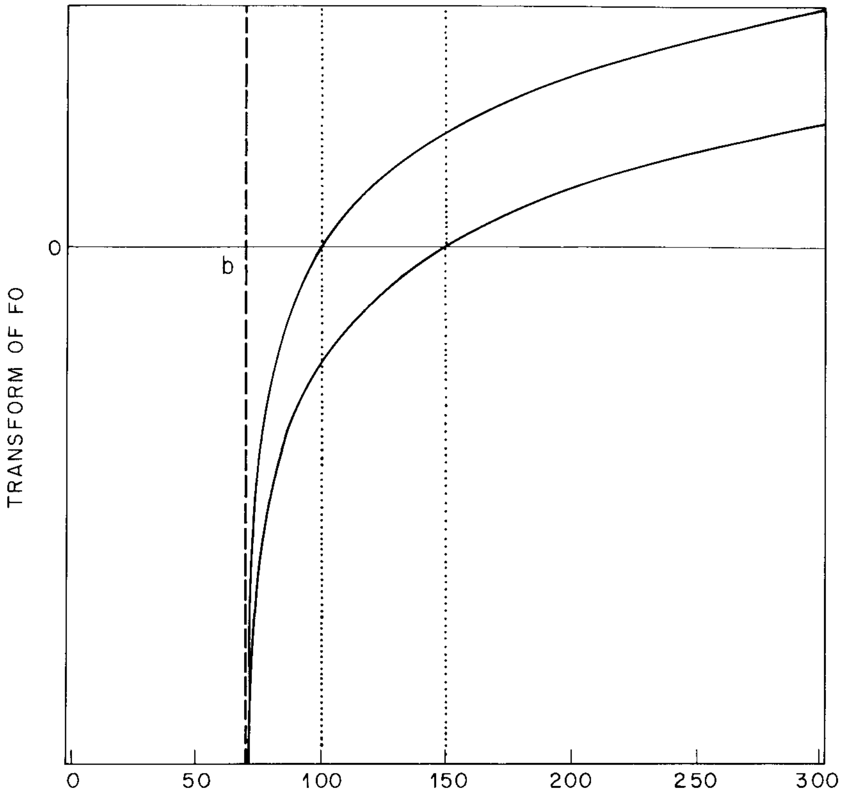


Figure 29
The relation of P, in Hz, to T(P) for two different choices of *r* in model 2. Values of P between *r* and *b* are transformed into values of T(P) between 0 and $-\infty$. Values of P between *r* and $+\infty$ are transformed into values of T(P) ranging from 0 to $+\infty$.

Table 13
Error measures for model 2

	DS data	AB data
MYL	2.7	5.2
DWS	3.5	5.3
JBP	7.0	7.2
KXG		7.7

A case in point is the model proposed in Liberman and Pierrehumbert (1979) and Pierrehumbert (1980). This model was developed to account for the structure of the AB/BA data and for the informal observation of downtrending F0 patterns in normal speech, as exemplified in figures 7 and 8. It includes a baseline that declines linearly over the duration of a phrase, with starting and ending values fixed, and a scaling function for F0 peaks of $(P_i - b_i)/b_i$, where P_i is the peak value in Hz and b_i is the value of the baseline at the point in the phrase where P_i occurs. It lacks lowering of final peaks, and it lacks any parameter like r to increase the asymptote of downstepping contours with increasing pitch range. As a result, it fits the downstep data very badly indeed, and the nature of its failures, as seen in the pattern of the residuals, pointed us in the direction of the models we have presented here.

5.3.1 Down with Rises and Falls The possibility remains that we are modeling the wrong sort of measurements. Our approach is in the general tradition of the American structuralists, who imagined that intonation should be described as a sequence of pitch level targets (or "contour points," to use Pike's phrase (1945)) connected by appropriate transitions. An opposing tradition of at least equal antiquity supposes that configurational features, such as "scoop-like rise," are crucial. Like the target level theory, the configurational feature theory lends itself to many embodiments; a first-order hypothesis might be that we should be modeling measured F0 rises and falls, rather than F0 levels (as suggested in Bolinger (1958), Bailey (1970, 1971), Clark (1978), Ladd (1978)). Perhaps the problems of the imaginary r and its relation to P_0 would then vanish, exposed as artifacts of our fundamental misconception.

There is some uncertainty about exactly what rises or falls should be modeled. Those linguists who claim rises and falls to be primitive have not given any instructions for measuring them, so we must improvise. In the case of the Answer-Background data, we measured the F0 at the beginning, peak, and lowest point of the first and second pitch accents—these points seem like reasonable ones. In the case of the Downstep data, the situation is more complex, as the F0 contour shown in figure 13 suggests, and we are puzzled about how to give the rise-fall theory a fair test. For lack of an obvious alternative, we will assume a "stairstep" idealization of the contour, using our original measurements of the peak value on the accented syllables as estimates of the step levels, so that the configurational model concerns itself with the movement from one step to the next.

Table 14
Answer-Background experiment, R squared for various regressions

	AB order					BA order				
	Peaks	Differences		Ratios		Peaks	Differences		Ratios	
		Rises	Falls	Rises	Falls		Rises	Falls	Rises	Falls
MYL	.92	.81	.89	.60	.83	.88	.80	.83	.73	.75
DWS	.93	.66	.61	.42	.28	.90	.75	.79	.58	.53
JBP	.90	.62	.80	.23	.46	.93	.60	.88	.11	.56
KXG	.84	.49	.46	.18	.04	.88	.44	.79	.11	.22

For the AB data set, the configurational approach seems to fail badly. Figures 30 through 33 show scatter plots of rises against rises and falls against falls for the AB experiment data for subject JBP. In comparison with the patterns in figures 14 through 17, these data are quite unruly. The patterns for the other subjects are similar: table 14 sums up the relative coherence of the relations between peak and peak, rise and rise, and fall and fall, as measured by the R squared value from linear regression. The rises and falls are treated both as differences and as ratios in Hz. As figures 30 through 33 suggest, the lower correlations of the rises and falls, compared to the peak measurements, are not due to nonlinear relations, but more simply to unsystematic ones. It appears that rises and falls are not being as carefully controlled as relative peak levels are.

In the case of the Downstep data, the configurational theory, in its simplest form, suggests that the differences or ratios between successive step levels should be the primitive measures. The ratio version is equivalent to model 1 with no reference line r, and thus fails badly. The difference version is no better, since the successive differences are systematically unequal. It may be argued that the configurational primitives should be the step-level ratios after r has been subtracted, with r rising with pitch range in an appropriate way. This model is identical to our model 1 and thus will fit quite well, but nothing about it seems especially attuned to the spirit of configurational theories.

5.4 The Need for New Evidence

As observed above, model 1 fits the data pretty well. There is a glimmer of a problem in the residuals for two subjects, but the problem seems to involve the aspect of the model that has the least theoretical coherence, and the effect is too uncertain to merit further discussion.

The experiments analyzed here were designed to clarify some questions

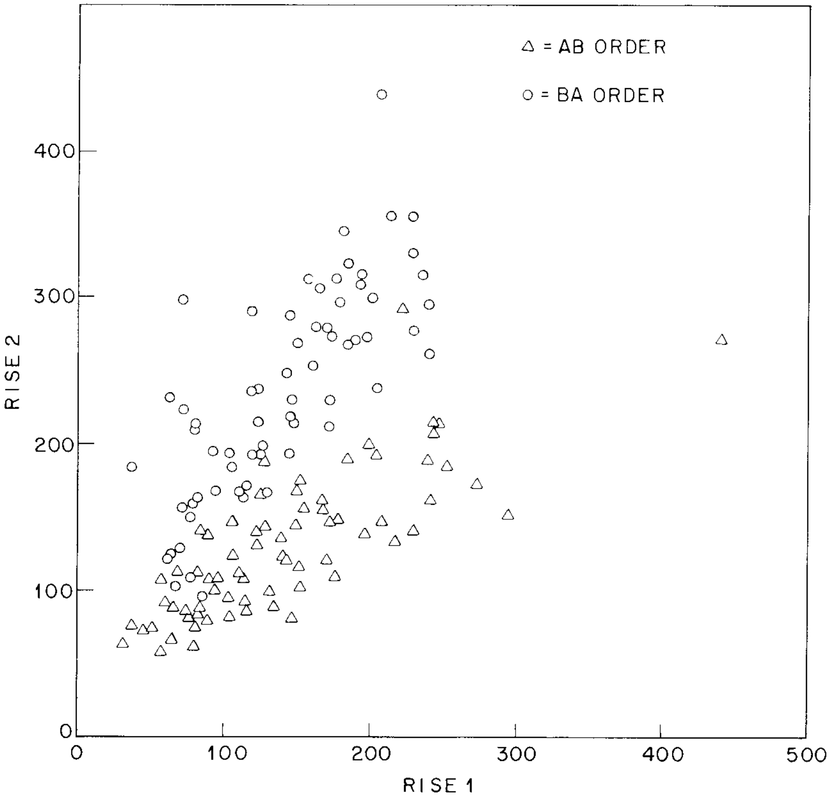


Figure 30
The size of the first peak rise plotted against the size of the second peak rise, for the AB and BA contours for subject JBP.

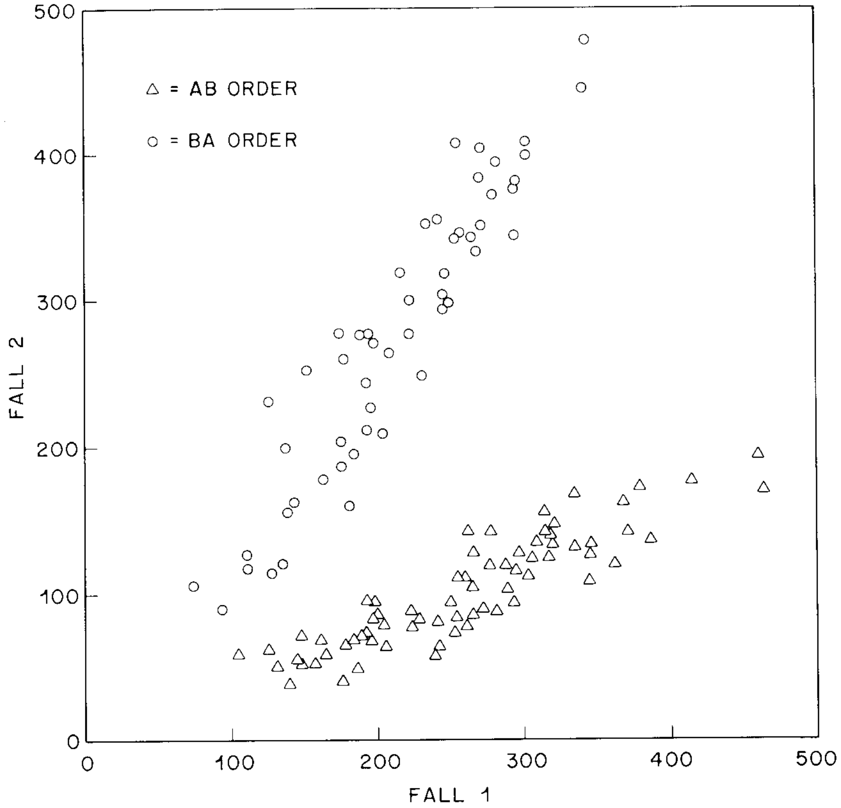


Figure 31
The size of the fall from the first peak plotted against the size of the fall from the second peak, for the AB and BA contours for subject JBP.

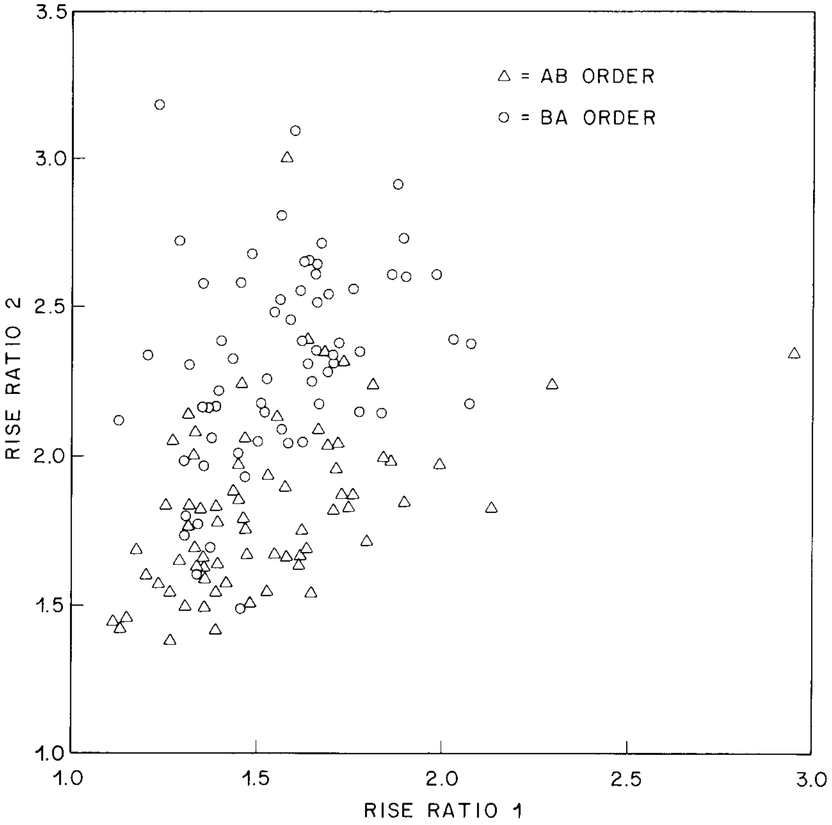


Figure 32
The ratio of the beginning and ending values of the first peak rise, plotted against the ratio of the beginning and ending values of the second peak rise, for the AB and BA contours for subject JBP.

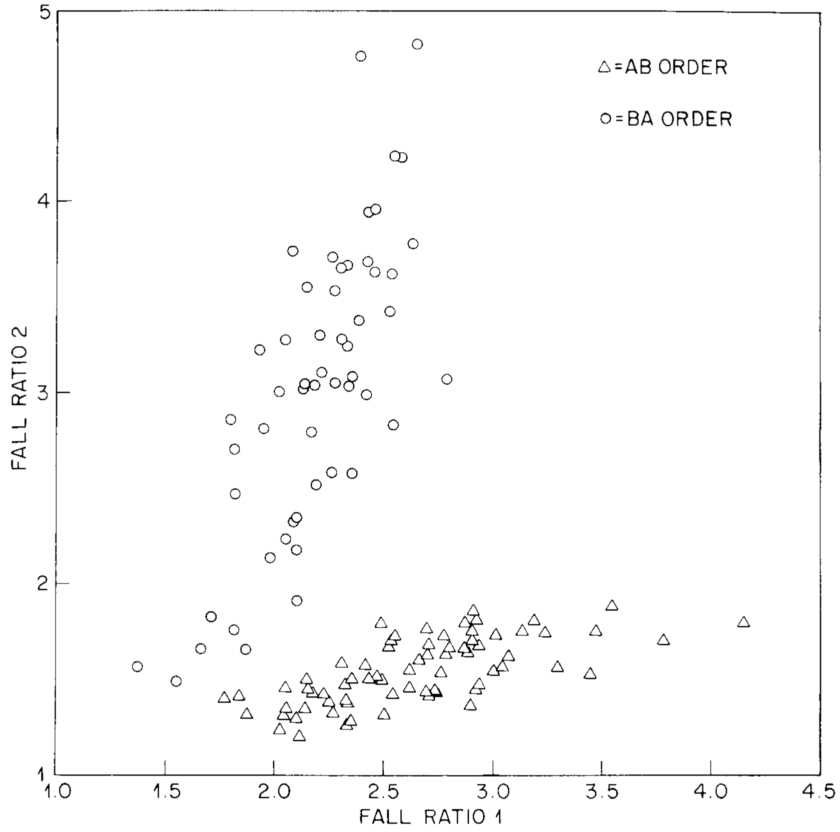


Figure 33
The ratio of the beginning and ending values of the first peak fall, plotted against the ratio of the beginning and ending values of the second peak fall, for the AB and BA contours for subject JBP.

that arose in the course of a broader investigation into English intonation. We claim to have learned something about the phonetic parameters that underlie English F0 contours, and about the relation between these parameters and some aspects of phonological descriptions. To learn more, we must test our model, suitably generalized, against a wider variety of cases.

A number of questions suggest themselves (see section 6.1 for an introduction to some terms): What is the behavior of English L* or L* + H accents under prominence and pitch range variation? What happens in cases where the phrase accent is H, and in particular, does "final lowering" apply? How does the "Answer-Background" relation behave when the two phrases each contain more than one pitch accent, or when there are multiple "Answer" or "Background" phrases? What structures can exist among sequences of downstepping phrases? Can r really vary independently of $P_0 - r$, and if so, how can the speaker's intent be recovered from the F0 contour? Do "upstepping" sequences exist, and if so, how do they behave? Our hypothesized phonetic parameter r is associated with phrase-sized domains—are there phonological entities that characterize such intonational domains? For instance, can the phenomena associated with the so-called phrase accent be profitably viewed in this light?

These questions, like the questions we have tried to answer here, confront the phonetics and phonology of intonation simultaneously.

6. Discussion

6.1 Some Remarks on English Tonology

Certain properties of the phonology of the intonation patterns we studied were integral to our treatment of their phonetic realization. If our approach to the phonetics of intonation is to extend, these crucial properties need to hold for the phonology of English intonation in general. This section sketches a theory of English tonology, proposed in Pierrehumbert (1980), that meets this requirement. Many of the details of this theory are not crucial to our treatment of the experiments described here; our aim is to suggest that our approach to English tonology can be coherently instantiated in a theory that accounts for a broad range of English intonation patterns.

Our most fundamental assumption is that tunes can be decomposed into a sequence of elements that are aligned with the text. These are pitch accents, which fall on some but not necessarily all stressed syllables, and additional tonal elements at the edges of the phrase. Decomposing the tune in this way provides the basis for realization rules like (5b), which compute the value for a tone by looking back to a preceding one. Our phonetic data provided some support for the claim that the primitive elements that make up the tune are tone levels rather than tone changes. We accept Bolinger's point that a theory with four phonologically distinct tone levels has an excessive number of degrees of freedom, given that local prominence and overall pitch range also introduce variation into the system. Reducing the tonal inventory to two appears to be the most direct way of alleviating this problem. Reduction to one tone (e.g., only upward-going excursions from a phrasal baseline or reference line, of whatever shape) would leave us unable to represent the full range of English intonation patterns.

In the grammar of English intonation proposed in Pierrehumbert (1980), all tunes are represented as strings of Low and High tones. The pitch accents consist of a single tone, or of a pair of tones, one of which is marked with * for alignment with the accented syllable. There are additional tones at the margins of the phrase: boundary tones that determine the F0 at onset and offset, and a "phrase accent" (as proposed for Swedish in Bruce (1977)), which consists of a single tone that occurs right after the nuclear pitch accent.

Under this account, the peak accents studied in the first experiment are H accents. In the B configuration, there is an L phrase accent followed by an H boundary tone; the A configuration ends with an L phrase accent and an L boundary tone.

The step accents of the second experiment are represented as a pair of tones, H + L. When one step accent follows another, its H tone is lowered by equation (5b). This downstep rule puts the H on the same level as the L of the preceding pitch accent, so that a staircase pattern results. The fall at the end of the phrase is due to an L phrase accent and L boundary tone; this is why the terminal values in the Downstep experiment are comparable to the terminal values for A configurations appearing in the BA order.

Many descriptions of African tone languages (see the surveys in Welmers (1973), Anderson (1978) use a downstep rule to generate a potentially unbounded number of phonetic tonal levels from an underlyingly two-tone system. The classic form of such a rule lowers H after a preceding HL sequence. If the downstepped H is on the same level as a preceding L, the system is said to have total downstep, while partial downstep leaves the H somewhat higher than a preceding L.

According to Pierrehumbert (1980), the English downstep rule resembles well-known African examples in lowering H after L. It differs in being triggered specifically by the L + H and H + L pitch accents. In the African examples, tones are not organized into pitch accents, and the

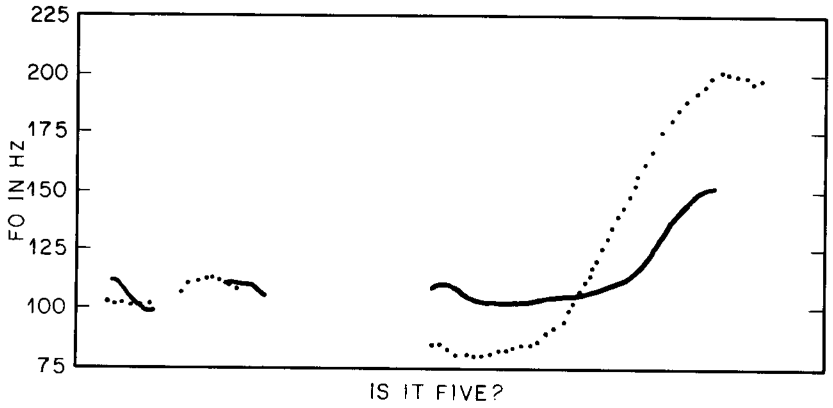


Figure 34
A pair of question contours with different degrees of emphasis on the questioned element. The L* pitch accent is realized at a lower level when the emphasis is increased.

downstep rule applies across the board to given tonal sequences. Restricting the English downstep rule to being triggered by the H + L and L + H accents is necessary to avoid overapplication. In particular, we assume that downstep does not apply to the second peak in the AB and BA sequences, in spite of the L tone(s) intervening between the first and second peaks. If the second peak were both downstepped and subjected to final lowering, it would have been lower than it was observed to be. An attempt to fit the data on this hypothesis would not have succeeded as well as our model 1. We presume that downstep is blocked in the AB and BA contours because the tones are not organized into the requisite pitch accents, and because of the intervening intonational phrase boundary.

When L and H tones are not linked in a two-tone accent, their behavior exhibits two interesting contrasts. We have noted that H* accents scale upward as the local prominence increases. This is not normally the case for L* accents. Increasing the local prominence on an L* accent causes it to scale downward, other things being equal.[10] This point is illustrated in figure 34. Second, the downward scaling is bounded in a way in which upward scaling is not. L* accents, as well as the L phrase accent and L boundary tone studied in the first experiment, appear to occupy a rather restricted region in the lower part of the F0 range. As prominence increases, the L* accents approach the baseline, but there is clear evidence of saturation. We have seen no evidence of saturation as H accents are raised

under increased prominence, although there must in principle be a ceiling on the F0 value a speaker can achieve.

One might draw an analogy between the floor and the ceiling on F0 and the floors and ceilings of everyday life. Jumping high can raise the issue of hitting the ceiling. However, our normal operations are carried out in the lower part of the room, with our feet on or near the floor. So the floor has more of a role in the planning of motor activities than the ceiling does.

The contrasts in the behavior of H and L tones follow from a transform like $\log((P - b)/(r - b))$, if we assume that the transformed value of a free L (that is, an L not in a two-tone accent) is the negative of the transformed value of an equally prominent H. Since we have not collected systematic evidence on the scaling of L tones, our suggestion that their scaling is symmetric to that of H tones is speculative. Another important question that we have left unanswered is how two-tone accents behave under changes in prominence.

6.2 Some Physiological Issues

Interesting questions arise about the physiological correlates of the patterns we have observed in our data and attempted to capture in our models. In particular, our posited "reference level," and the phenomenon of final lowering, call out for physiological explication. Some aspects of the system could also in principle be rooted in the perception of F0 patterns, but it is our view that the physiology of production is a more likely source.

Final lowering and variations in the reference level might come about through variations in pulmonary control and/or the rest state of the larynx. In particular, it is tempting to attribute phrase-level effects to subglottal pressure, as Collier (1975) proposes. However, our perusal of data on subglottal pressure and F0 available in the literature (Collier (1974), Atkinson (1973)) suggests that subglottal pressure does not vary with a time course that would permit it to control final lowering.

Independent of the mechanism for these effects, there is the question of whether they are under phonological control or not. It is not hard to think of experiments to investigate this question. If, for example, final lowering applied to the last (nuclear) accent regardless of how many postnuclear syllables intervened before the phrase boundary, we would conclude that it was under phonological control. If, on the other hand, it occurred in a fixed timing relation to the end of the utterance, it would seem more likely to be a phonologically irrelevant consequence of the way utterances are ended.

It is possible to ask the same question with regard to the reference level.

Under one kind of phonological control of the reference level, each reference level setting would be associated with a phonological domain, perhaps the intonational phrase. There might perhaps be rules relating reference levels of different phrases, just as downstep relates the values of different tones. On the other hand, our reference level might better be thought of as a parameter at best only loosely coupled to linguistic structures, on the model of eyebrow height or even heart rate.

6.3 The Problem of Preplanning

Recently, there has been considerable interest in the question of phrasal preplanning and in the implications of the F0 declination effect for this question. In order to consider these issues sensibly, it is helpful to distinguish between what one might call "hard" and "soft" preplanning. By "hard" preplanning we mean processing that is an essential part of intending to say something and that normally needs to be accomplished before executing that intention, on pain of dysfluency. By "soft" preplanning we mean the sorts of preparation that a speaker may freely choose to make, out of rational calculation, ritual observance, or any other cause, and that might well be omitted for a linguistically equivalent utterance under other circumstances. An extreme example of "soft" preplanning might be preparing to duck before uttering an insult. This distinction probably names the ends of a continuum, but it is useful nonetheless.

Those who find the terms "hard" and "soft" unpleasantly evaluative may prefer to distinguish different levels of planning. For instance, there is doubtless a correlation between the duration of a planned journey and the weight of the baggage carried at its start; this correlation provides evidence of "preplanning" and is perceptually potent, in that we may infer our acquaintances' plans from the quantity of their impedimenta. But the psychology of locomotor control concerns itself with plans at quite a different level.

Under appropriate circumstances, "soft" preplanning of some aspects of F0 control must surely occur. Obviously speakers can choose a pitch range at will, sometimes speakers know in advance about how long a phrase will be, and there is nothing to stop them from using higher pitch ranges for longer phrases if they think this will suit their purposes. The desire to have plenty of space for stepping down might urge such a strategy; on the other hand, the desire to save one's breath for the long pull might well urge the opposite. In our Downstep experiment, the subjects used neither strategy to any very large extent, but little significance can be attached to this result.

Any observed association of pitch range choice with phrase length is a natural candidate for explanation as "soft" preplanning. We would expect such effects to be erratic and generally small compared to other influences, and we are confident that speech lacking such correlations does not sound in any way abnormal, since any random stretch of natural talking provides an experimental test.

In general, we see no evidence that the F0 implementation for an entire phrase is necessarily laid out before speaking begins, even when the phrase is known in advance and fluently produced. All of our measurements can be modeled quite well on a left-to-right, plan-as-you-go basis. Negative evidence of this kind is logically rather weak—perhaps we simply looked at the wrong contours or at the wrong measurements. However, we think that the phenomena sometimes argued to provide evidence for preplanning of F0 implementation (e.g., by Cooper and Sorensen (1981)) are easily handled by the type of theory we have proposed, one that lacks any sort of "hard" preplanning of whole phrases.

There are two intonational descriptions in the literature that address the question of preplanning in intonational realization, those of Cooper and Sorensen (1981) and Thorsen (1980b). Cooper and Sorensen present a description of English F0 peak values in which the role of preplanning is emphasized. Two consequences of preplanning are proposed: an increase in first peak height with phrase length, and a phrasally-based computation of medial peak values, by the *Topline rule*. In Cooper and Sorensen's data, the height of an initial F0 peak increased with sentence length. They do not handle this effect formally, since it is unclear whether the relevant measure of length is phonological, syntactic, or semantic. The Topline rule computes medial peak values as a function of first and last peak values. An implication of this rule is that the value of a noninitial peak cannot be computed without knowing the final peak. We have three reasons for rejecting this conclusion. First, the outputs of the Topline rule resemble downstepped contours, which we have been able to model without assuming preplanning. Second, Cooper and Sorensen do not compare their model to models that do not rely on the final peak, so that the need for the final peak in predicting earlier ones is uncertain. Third, serious errors in their statistical analysis leave their claim that their model fits the data unsupported. These points are developed at greater length in Pierrehumbert and Liberman (1982).

A more serious challenge to the idea that intonational realization can be handled by local rules is the work of Thorsen on Danish. Thorsen has studied the realization of Danish pitch accents, which we would analyze as

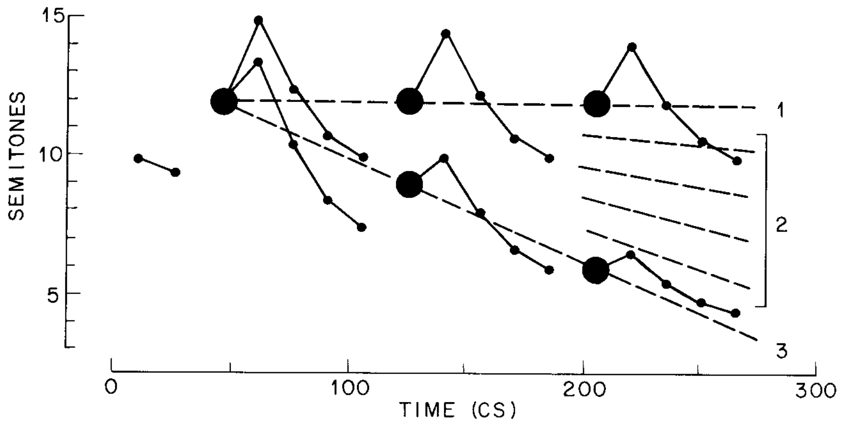
L* + H, in utterances involving one to eight accents. Her materials include terminal and nonterminal declarative phrases, as well as interrogatives. Figure 35 summarizes her conclusions about the regularities of Danish F0 contours. In the figure, the peaks fall gradually from the beginning to the end of an utterance. The slope of the fall depends on the sentence type: it is steepest for terminal declaratives and least steep for interrogatives, with nonterminal declaratives falling in between. Thorsen takes the downslope to arise from declination, and asserts that the slope of declination depends on sentence type.

In order to interpret Thorsen's data in terms of a model like ours, it is necessary to assume that the downslope arises from a downstep rule. This assumption seems plausible, since Danish has the classic context for downstep, alternating Hs and Ls. The slope differences would, in our view, most plausibly arise from different choices of relation between r and P_0. The steepest contours would have the lowest r, and raising r would make the interrogative and nonterminal declarative contours less steep, for the same P_0 by raising the asymptote for stepping. On this account, one might expect P_0 to run higher in sentences with higher r, since there is no a priori reason for initial prominence to be inversely correlated with r. Thorsen's data show a tendency for P_0 to be higher in sentences with less overall downslope. However, $P_0 - r$ evidently decreases as r is raised. This effect might represent a ceiling effect on choice of P_0 in a normal speaking voice.

An alternative description of the slope difference in Thorsen's data would rely on varying the step factor s with sentence type.

Given these assumptions, Thorsen's data may be compared to our data on step accents in sentences of varying length. Figure 36 shows data on terminal declaratives for Thorsen herself, whose data was the best behaved of any of the subjects'. As in our data, the last peak is lowered, so that its value falls below that for nonfinal peaks at the same serial position. However, in the longer sentences it appears that lowering is not confined to the last peak.

There are at least two interpretations for this fact. One possibility is that the physiological process that is responsible for final lowering in our data—whatever it may be—appears further from the end in Thorsen's longest sentences, because her seven- and eight-accent sentences taxed speakers in a way that our five-accent sentences did not. Another possibility is limited look-ahead in the principles that determine F0 target levels. This would require relaxing our position somewhat, but it is still very different from a theory with global realization rules.



A model for the course of F0 in short sentences in ASC Danish. 1: syntactically unmarked questions, 2: interrogative sentences with word order inversion and/or interrogative particle and nonfinal periods (variable), 3; declarative sentences. The large dots represent stressed syllables, the small dots unstressed ones. The full lines represent the F0 pattern associated with stress groups, and the broken lines denote the intonation contours. Zero on the logarithmic frequency scale corresponds to 100 Hz.

Figure 35
A summary of Thorsen's conclusions about the regularities of Danish F0 contours, reproduced from Thorsen (1980a).

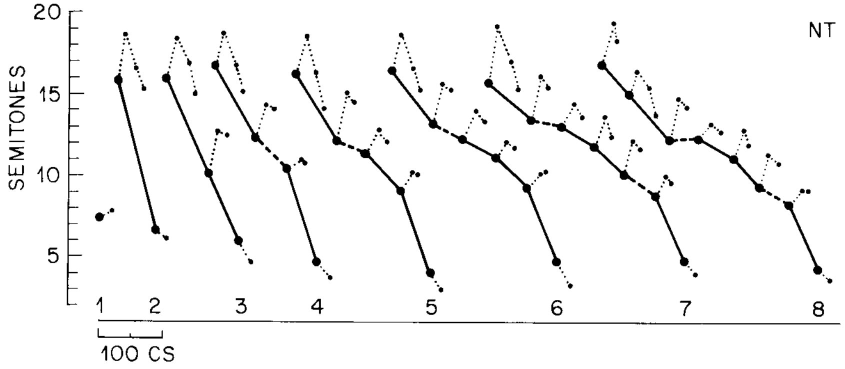


Figure 36
Data on terminal declaratives taken from Thorsen (1980a). The subject is Thorsen herself. Large dots represent stressed syllables and small dots represent unstressed syllables. Thus, dotted lines connect values for each stress group, while the solid and dashed lines trace out the intonation contour, defined by Thorsen as the sequence of values for stressed syllables. Zero on the logarithmic frequency scale corresponds to 100 Hz.

6.4 Comparison to Other Work on F0 Modeling

6.4.1 The Great Declination Debate We noted in the introduction that a number of authors have reported a gradual downtrend in F0, which is often given the name *declination*. Other authors, for example Lieberman and Tseng (1980), have contested the existence of such an effect, while Umeda (1982) regards the effect as "situation dependent." Our work brings the methodology used on all sides of this debate into question. Essentially all of the participants attempt to measure declination on material that is not analyzed (and thus not controlled) in terms of phrasing, stress pattern, and tune; and none of them distinguishes among the various plausible contributions to F0 trends that we have discussed, i.e., global decay, boundary effects such as initial raising or final lowering, local tonal downstep, correlations of stress pattern or semantic/pragmatic prominence with order, and the interaction of all of these with changes in pitch range and initial peak height. It follows from what we have learned here that such measurements are uninterpretable.

This point can be brought out by considering the surface form of the intonation patterns studied in this paper. The AB patterns have a surface downtrend through the peaks, but a line fit through the lowest points would rise, since the A accent ends in an L boundary tone, but the B accent does not. In the BA patterns, the lowest point in the second phrase is lower than the lowest point in the first. However, here the second peak is usually higher than the first. We have seen, however, that the phrase-final lowering effect is the same in both cases. This conclusion is not evident from measuring the peak or valley trends in either contour; it is reached only by separating the phrase-level effects from other factors controlling F0 implementation, on the basis of a theory about what these factors are.

The same point emerges again in considering the downstepped patterns. These patterns, unlike the AB/BA patterns, exhibit stereotypical declination, in that both the local maxima and the local minima in the contour are monotonically decreasing. However, most of the downtrend should be attributed to the choice of pitch accent type, rather than to a phrase-level effect on the F0 contour. It is only when the downstepping due to the pitch accents is taken out that the phrase-level effect (i.e., final lowering) turns out to be same as that observed in the AB and BA contours.

Recall also that models like ours yield differing rates of apparent "declination" (for a given choice of phrasing, stress pattern, and tune) as a function of pitch range and initial peak value. Thus, reports of overall slope differences among intonation types, such as Waibel (1979) and Thorsen (1980b), may reflect control of parameters that influence observed downtrends only indirectly.

6.4.2 Some Log Models Among authors who have tried to be explicit about how F0 contours are scaled, log transforms are popular. 'tHart and Cohen (1973) propose a model of Dutch intonation in which "hat" patterns are added onto a declining baseline in the log domain. In Lea's model for English (1973, 1979), utterance, phrase, and stressed-syllable components are added together in the same way. Such models make wrong predictions because they lack any counterpart to our reference level, which changes with overall pitch range while leaving the baseline (seen in the final L tones) invariant.

6.4.3 Fujisaki's Model Fujisaki[11] has suggested a very interesting model of F0 realization, following an approach originally laid out by Öhman (1967). Fujisaki's proposal models F0 contours by summing (in the log Hz domain) three independent contributions: (1) a constant value, *Fmin*, which "indicates the lower limit of fundamental frequency below which vocal fold vibration cannot be sustained"; (2) a *baseline component*, which is a phrase-sized falling pattern; and (3) an *accent component*, which implements local excursions due to pitch accents.

Fujisaki explains that "F0-contours of words and sentences are generally characterized by a gradual declination from the onset towards the end of an utterance, superposed by local humps corresponding to word accent." Both the baseline component and the accent component are modeled as a succession of discrete levels, "smoothed separately by the low-pass characteristics of their respective control mechanisms," which are approximated as critically damped second-order linear systems. Each baseline unit and each accent unit has an amplitude, a start time, and an end time.

Fujisaki's model treats accents and baselines differently. After smoothing, each accent unit rises from zero, at its start, toward its specified amplitude level, which it approaches asymptotically, and then falls asymptotically to zero after its end. Accent units are assumed not to overlap in the input—an accent's start time cannot precede a previous accent's end. Baseline units, on the other hand, all have their ends at the end of the utterance (or perhaps the "breath group"), so that successive such units ride on the tails of their predecessors. The smoothing of baseline units is also different, so that their contribution rises towards the specified amplitude and then falls gradually towards zero. At the end time of a baseline unit, a negative copy of its onset is added in. The start time of baseline units is generally placed about 200 msec. before the beginning of the corresponding phrase, so that the baseline peak lines up approximately with the start of the phrase. The placement of baseline units' end time with respect to the end of the utterance seems to vary somewhat, so that a variable amount of the negative-going region of the baseline component plays a role in the modeling.

We see two main difficulties with Fujisaki's model, which put it in qualitative disagreement with our data.

First, the negative-going region of the baseline component, which occurs after its end time, is an unfortunate feature. The amplitude of this negative-going region is the same as the amplitude of the positive-going part of the baseline component it arises from; and if more than one phrase exists in the utterance, then the negative-going tails of these phrase superpose. If the placement of the baseline component end times is held constant, and if these negative-going regions play any role in the modeling, then the model predicts that higher pitch range sentences will have lower final values, if they are long enough for the positive-going baseline component to have damped out. Under the same assumption, the model also predicts that subdividing a single intonational phrase into two phrases, thus adding up two baseline components, will lower the low value at the end of the utterance. Both of these predictions are false, at least for English.

In the cited references, detailed fits are given for three sentences. For these cases, the negative-going part of the baseline component plays no significant role, since the baseline end times are placed essentially at the end of the utterances. Figures are also provided showing the model's fit to four isolated words; here the nature of the baseline component is not specified, but it seems that in at least some of these examples, the post-end-time part of the baseline component has been used in order to get the F0 low enough at the end. This difference (use of the post-end-time part of the baseline for short utterances, no use for longer ones) seems to arise because the relatively slow time-constant of the baseline control mechanism (about 300 msec.) leaves the baseline component quite high for these single words, which are about 400 msec. long, while allowing it to decay to a low value for the multiword phrases, which are all longer than one second.

Our own experiments with Fujisaki's model suggest that this problem is a general one. In real life, utterance-final low tones are essentially invariant, and to make this work out in the model, the location of the baseline component's end time must be earlier in short low-ending utterances than in long ones.

A second difficulty with Fujisaki's model arises in trying to represent the

effects of pitch range change. With respect to our downstep data, for instance, his model lacks any effective mechanism to raise the downstep asymptote with pitch range. Because his baseline component decays to zero, increasing the baseline amplitude will not effectively raise the level two seconds or so into the phrase. Raising his parameter Fmin will make it nearly impossible to get the final low values to come out invariant; at best, just the right amount of the negative-going tail of the baseline component will have to be accessed in order to bring the final F0 values down in compensation.

We have spent so much time on Fujisaki's model because we think so highly of it. It attempts to model the entire time course of F0 patterns on the basis of a small number of linguistically motivated parameters, which are given plausible physiological interpretations. There can be little doubt that an appropriately modified form of this approach can be made to fit our data; however, they are incompatible with the model as Fujisaki presents it.

7. Some General Issues in Phonology and Phonetics

We wish to present our discussion of intonational implementation in English as an example of a more general approach to investigation of the sound of language.

The phonetic implementation of phonological representations is responsible for much of a language's sound pattern; this is true both for impressionistic and for instrumental observations of such sound patterns. Any treatment of the sound pattern of a language implies a particular division of labor between the principles that determine the distribution of phonological entities and the principles that govern the relationship of those entities to our observations. Implementation systems are neither trivial nor obvious; it follows that the correct division of labor between phonological representation and phonetic implementation is not obvious either. Better understanding of the relation between representation and implementation should lead to better theories of both.

The raw material of both phonology and phonetics is observation of patterns in the sound of entities like syllables, words, and phrases. Here are three such observations about English: (1) the indefinite article is realized as /ei/ or schwa before consonant-initial words, but as /æn/ (or related reduced forms) before vowel-initial words; (2) the aspiration of voiceless stops is longer before high vowels than before low vowels; (3) in American speech, a silent closure is often inserted between tautosyllabic nasals and voiceless fricatives, so that *tense*, for example, becomes rather like *tents*;

according to Fourakis (1980), this does not occur in South African English.

Observation (1) is usually considered to be a fact about English morphology. Could it instead be a fact about phonetic implementation? This seems unlikely, for a number of reasons. First, the variation in question is restricted to a single morpheme, and a description of the environment in purely phonological terms would be extremely ad hoc, at best. Therefore, an implementational treatment would involve lexical conditioning of phonetic implementation. While such conditioning may not be entirely avoidable, it seems to be the exception rather than the rule. Second, the observed sound pattern is exactly what is expected if the phonological representation contains an /n/ in one case and not in the other. Independently plausible rules and processes of reduction will then give the observed range of variants in each case.

Observation (2) can probably be explained on physical grounds, as Ohala (1974) argues. When the supraglottal impedance is greater, as it is for high vowels, vocal cord adduction must be further advanced for vocal cord oscillation to begin. Therefore, writing a phonological rule to modify aspiration duration across the appropriate environments would be a category mistake.

Observation (3) is more troublesome than (1) or (2). As Fourakis observes, the fact that the pattern is dialect-specific argues strongly against an explanation phrased entirely in terms of physiology and physics, such as the one suggested by Ohala. Fourakis proposes a rule of stop epenthesis, appropriately conditioned. However, Fourakis also presents measurements clearly showing that such "epenthetic stops" are systematically different from underlying stops in the same environment, being shorter by about 25% in the case of *tense* than the corresponding region of *tents*. On this account, the phonetic realization process must somehow know whether the stop is underlying or epenthetic. Fourakis suggests that the variation may be conditioned by a difference in syllable structure. It seems simpler to leave the phonological representations as /tens/ and /tents/ (or whatever featuro-syllabic translations of these one prefers) and let the "epenthesis" be accomplished by dialect-specific patterns of timing control for the velum, larynx, and tongue in the implementation of such representations. The case merits deeper study—our present point is simply that the implementational treatment of such epenthesis phenomena is a plausible one.

It has been our experience that cases of "allophonic variation" often turn out to have properties like those of observation (3). This leads us to suspect that a correct division of labor between phonological representation and phonetic implementation will leave the output of the phonology rather more abstract than it is usually assumed to be.

7.1 Favorable Consequences of Enhanced Status for Phonetic Implementation

Better understanding of phonetic implementation removes from phonology the burden of representing those sound patterns for which its natural descriptive mechanisms are inappropriate. As a result, the theory of phonological features and rules should be clarified and simplified.

For instance, the description of English intonation in Pierrehumbert (1980) is broader in coverage and simultaneously simpler in conception than the treatment given in Liberman (1975). The foundation stone of the improved edifice is the reduction of the tonal inventory to H and L, from the four-way contrast in Liberman. This reduction depends crucially on a set of nontrivial implementation rules for tonal sequences. Most of the details of individual F0 contours, naively related to local tone levels by Liberman, depend on the operation of the implementation rules. This paper revises those rules considerably, in the interests of empirical adequacy, but retains the phonological advantages of Pierrehumbert's treatment without change.

In another example, Prince (1980) argues that the feature of overlength should not figure in the phonemic inventory of Estonian, despite its central role in the sound patterns of that language. Instead, overlong vowels should be treated as an aspect of the phonetic realization of monosyllabic feet. His analysis provides an improved account of the ways in which "the three-way contrast, and in particular the distribution of overlength, is richly and curiously connected with patterns of morphology, syllable structure, and stress." As a result, an unusual three-way length contrast is plausibly reduced to the interplay of foot structure with the more familiar two-way length contrast. These gains follow from a more abstract conception of the relation between phonological categories and the traditional description of the phonetic surface of the language. Prince (1980) shows that one need not do independent phonetic research in order to demonstrate the benefits of this conception. To quote from his conclusion (p. 559): "A possible (if uninvited) reading of Liberman and Prince (1977) would be that it offers—merely—a pleasant new algorithm for distributing the familiar phonetic properties of stress (0, 1, 2, 3, . . .). A deeper possibility is that metrical theory involves a fundamental revision in the notion of phonetic representation: it changes our ideas about what the 'familiar phonetic properties' are."

For all types of phonological mechanism, the set of needed rules can be made more restrictive, without loss of generalization, if the implementation system is appropriately extended. For instance, in the autosegmental framework Clements and Ford (1979) propose a set of rules for Kikuyu downstep that move a downstep marker (analyzed as a floating L tone) over arbitrarily many adjacent L tones, all of which are changed to H. If this movement results in two adjacent downstep markers, both are deleted, although an earlier rule deletes only one of a pair of adjacent downstep markers (ar:sing from another source). In Prince and Liberman (1982), it is shown that a simple extension of the Clements-Ford system for the realization of downstep permits the downstep movement, the block raising, and the downstep deletion all to be eliminated.

7.2 Belling the Cat: What Can't Implementation Rules Do?

Lacking further constraint, we could in principle construct a system of "phonetic implementation" that directly interprets the level of logical form, thus simplifying to vacuity the theories of syntax, morphology, and phonology. Conscience and good taste restrain us, but we cannot expect the general run of infants and phonologists to share our sensibilities. Once we have granted that phonetic implementation is nontrivial, how can its appropriate domain of application be determined?

Putting it briefly, we don't know. Obviously enough, the answer is to be found in appropriate restriction of phonological features and rules, and of phonetic parameters and control strategies. We have one conjecture to offer about phonetic control strategies; it happens to be true of the implementation rules we have proposed, but this is partly because we have had some form of the conjecture in mind. The main purpose of this conjecture, aside from the usual provocation of attempts at refutation, is to provide a crude example of the type of restriction one might aim for.

We assume that various phonological objects have phonetic parameters associated with them. For instance, in our model the parameter r is a property of an intonational phrase, while P is a property of a pitch accent or perhaps of one of its constituent tones. The set of phonetic parameters presumably has about the same degree of universality as the set of phonological features. When a parameter value has "paralinguistic" meaning, as in the case of pitch range, this can best be modeled by the interpretation of free selection among the available alternatives, as in the case of lexical choice. Other aspects of parameter determination are thought of as the expression, in the current context, of the intrinsic phonological content of the object in question.

Now comes the conjectural part: we insist that the computation of any parameter of object Y_i can only depend on the "accessible" properties of Y_i and Y_{i-1}, where Y_{i-1} means the immediately previous object of the same type (if any). Thus, pitch accent can look back to previous pitch accent, phrase to previous phrase, etc. "Accessible" is taken to include a small set of intrinsic properties of the objects and any relations between them; it is intended to exclude any properties of the subconstituents of these objects.

A restriction of this type has a certain functional value, for both speakers and hearers; speakers can get on with the task at hand without knowing all the details of what follows, while hearers can in principle complete the phonetic processing of what they have heard up to any given point in the stream of speech.

Many apparent instances of anticipatory effects are known. Unless our conjecture is wrong, all such cases must turn out to be explained either by feature spreading at the phonological level, by computation of some parameter of a higher-level constituent, by interpolation between one target and the next, or by temporal overlap in the realization of the segments in question.

7.3 Lexical Phonology and a Possible Division of Labor

According to the theory of lexical phonology, as described in Kiparsky (1982, 131), "the derivational and inflectional processes of a language can be organized in a series of levels. Each level is associated with a set of phonological rules for which it defines the domain of application... This establishes a basic division among phonological rules into those which are assigned to one or more levels in the lexicon, and those which operate after words have been combined into sentences in the syntax."

In such a framework, the minimally required set of postlexical rules would combine lexical representations into a well-formed phrase-level phonological structure. One reasonable account of the division of labor between phonological representations and their phonetic implementation would limit postlexical rules to such a minimal set and assign all other postlexical regularities to phonetic implementation. (Note that if we permit "floating" segments and/or empty structural positions to emerge from the lexicon, then a certain amount of phonological readjustment of the edges of words is still permitted, as in the case of French liaison.)

8. Conclusion

Like any piece of rational investigation, this paper deserves to be judged separately according to its contributions in descriptive particulars, in theoretical proposals, and in style of research.

Its descriptive particulars concern the phonetic parameters underlying English intonation and their role in relating abstract intonational categories to observed F0 contours. We have proposed to reinterpret F0 measurements in terms of a fixed "baseline," a "reference line" that increases with pitch range, and a lowering effect specific to the domain of (certain) final pitch accents; this approach yields F0 implementation rules that do not require whole-phrase preplanning.

Our theoretical proposals concern the general status of phonetic interpretation of phonological categories. We have suggested that nontrivial, (partly) language-particular phonetic interpretation should take over the function of most postlexical phonological rules. This suggestion may stand or fall quite independently of the intonational descriptions that engendered it.

If either our descriptions or our theories find favor with the reader, we hope that our style of research will be given due credit. Unfortunately, there are distressingly few researchers who give serious thought both to phonological descriptions and to phonetic measurements. We feel that hybrid research strategies, in which the methods of phonology and of phonetics are simultaneously exercised, will play a crucial role in improving theories of human speech and language.

Notes

1. Among others: Fujisaki (1981), Fujisaki, Hirose, and Ohta (1979), Fujisaki and Nagashima (1969), Fujisaki and Sudo (1971) for Japanese: Vaissiere (1971) for French; Thorsen (1980b) for Danish; Lea (1973, 1979), Liberman (1975), Maeda (1976), O'Shaughnessy (1976), Sternberg et al. (1980), Cooper and Sorensen (1981) for English.

2. As suggested in Trager and Smith (1951), Pike (1945), Liberman (1975).

3. See section 6.3. Of course, lack of long-range effects is not evidence against preplanning, but merely lack of evidence for it.

4. Previous studies of such patterns include Jackendoff (1972) and Liberman and Sag (1974).

5. The designations *A* and *B* for these configurations are taken from Jackendoff (1972). It is important to note that the B configuration is not the same as Bolinger's

(1958) B pitch accent, which has a different shape and can occur at any position in the intonation phrase.

6. Such effects are documented in Peterson and Barney (1952) and Lehiste and Peterson (1961). Lea (1973) reviews the literature on this subject.

7. Actually, some other patterns were included for some subjects, raising the number of cases in a block to 40 or 60.

8. We tried to avoid using scattered points that resulted from mistracking in noise or from vocal fry. In the first experiment, if any part of the final syllable was tracked consistently, we used the lowest value in the well-behaved sequence. In the second experiment, we followed the same rule, except that if a large part of the last syllable was lost to vocal fry, we did not record a measurement.

9. In writing this paper, we have tried to explain mathematical concepts that some readers may not be familiar with. We beg patience of mathematically sophisticated readers.

10. It is important to remember that increasing the overall pitch range raises an L tone of a given amount of prominence, by raising *r*. Since local and overall effects of increasing emphasis influence the value of L tones in opposite directions, the picture is qualitatively more complex than the picture presented by H tones.

11. Fujisaki and Sudo (1971), Fujisaki, Hirose, and Ohta (1979), Fujisaki (1981). Quotations in the following description are from Fujisaki (1981). These works deal specifically with Japanese, but are argued to apply to English as well.