Peak Vary, Endpoints Don't: Implications for Intonation Theory
Author(s): Suzanne Boyce and Lise Menn
Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society (1979), pp. 373-384

PEAKS VARY, ENDPOINTS DON'T: IMPLICATIONS FOR INTONATION THEORY

Suzanne Boyce, Massachusetts Institute of Technology
Lise Menn, Aphasia Research Center, Boston Veterans Administration Hospital

Phoneticians have occasionally remarked that the final fundamental frequency (F_o) value of intonation contours with terminal falls is generally constant over many different utterances by the same speaker. The phenomenon seems to be common knowledge, borne out by the personal observations of a number of researchers, but it has not been adequately studied or documented in the literature. Some reference to it may be found in instrumental work (Phillips 1970, Maeda 1976, Sorenson & Cooper 1978), but in none of these cases was it an important subject of discussion; it seems to have been regarded as an interesting, physiologially plausible, but incidental fact about intonation.

In this paper we show how we confirmed the presence of this phenomenon of end-point stability for the spontaneous speech of a larger set of speakers. We then show that this simple fact may have interesting implications for the theory of intonation.

Method: Sample. The corpus consisted of six parent-child conversations which were recorded by Jean Berko Gleason's research group (at Boston University) as part of a larger study of parent-child communication. The tapes were made in a laboratory playroom; three mother-child and three father-child conversations were analyzed. The children were between the ages of of 2:0 and 4:0[1]. These conversations are probably as spontaneous as can be obtained when the participants know that they are being taped.

Each sample consisted of about 100 successive utterances. Normally about 65 of these were uttered by the parent and the other 35 by the child. About half of the total output of each speaker consisted of falling-contour utterances which were judged to sound 'complete'; the rest was made up of rising contour sentences, questions with tag endings, and falling-contour sentences which were judged to sound unfinished. Most of the falling-contour sentences were declarative statements and wh-questions. Imperatives were found with a variety of contours.

Data: The data analyzed in this paper are measurements of the maximum F_o reached in each clause and the endpoint of each falling-contour sentence (see figs. 1a and 1b). We included one-word utterances as sentences; both interjections and nouns were frequently found in isolation, especially as answers to questions.

F_o was measured by a pitch extraction program developed by W. Henke of MIT. This program handles both parents' and children's voices with a high degree of accuracy (error being within approximately 5 Hz). Narrow-band spectrograms were used to supplement our data, and also used routinely to check the accuracy of the computer data in 5 to 10% of the cases. The degree of accuracy obtainable in spectrogram data is more variable but was estimated to be within 10 to 15 Hz.

With respect to measuring endpoints, Philip Lieberman and others have called attention to the problem of deciding exactly where in time an utterance ends, since it may fade into inaudibility, be embellished by a change of register, or be otherwise obscured. Since most of our utterances ended quite cleanly, this was not a problem for us; repeated measurements of points yielded consistent results. In addition, the pitch at the end of most falling contour utterances has a rather small rate of change, as though it were approaching an asymptote. Thus, most errors due to temporal impreciseness of the endpoint are small. If the speaker dropped into creaky voice (which happened very seldom in our sample, the value of the last few pitch periods of the normal speaking register was taken to represent endpoint frequency.

The peak values of fundamental frequency in sentences presented little problem in measurement, being louder and better defined than falling-contour endpoints.

Noise, of course, is typically a problem in work with natural speech. We had several kinds: traffic noises coming from outside the playroom were the greatest nuisance. Here, low-pass filtering enabled us to recover the F_o of all but the worst sections of discourse. We did have to discard some utterances when speakers moved too far from the microphones, and some points were lost to the usual crashes, bangs, and clanks of toys. A few endpoints were also lost when one speaker interrupted another. However, most of our sentences were quite usable. Moreover, parent-child conversations have significant advantages for discourse analysis which more than compensate for the toy-noises. First, the increased pitch range gives much more variance to work with than one would find in adult-adult conversation. This increases the likelihood that any systematic variation that we find is of sufficient magnitude to work with: for adults speaking to two-year olds, F_o ranges over at least an octave and a half, sometimes two, while for the usual reading voice study, F_o ranges over just about one octave. Second, as is well known, parents' speech to children is simpler in many ways than adult-adult conversation. Semantically it is more concrete and less involved with subtle shades of meaning; the syntax is also simpler, and a higher proportion of sentences is fully grammatical. We suspect that intonation, as well, will be more repetitive and less subtle than in adult-adult conversation, although it probably does carry a higher affect load.

Results: When we examined endpoint values of those sentences judged to sound 'finished' for each speaker, we found that they seemed to show remarkable stability of the terminal endpoint F_o. On the other hand, the F_o values for all clause peaks varied tremendously and without regard to sentence type. This initial impression was confirmed when we looked at the statistical distributions of the peaks and endpoints of each speaker. Standard deviations of 'finished' falling-contour sentences ranged from 3 to 14% of the (logarithmic) mean endpoint value. In musical terms, this is about one-half to two and one-half semitones. Peak F_o, on the other hand, had a standard deviation ranging from 22 to 39%, or from three and three-fourths to six and one-half semitones from the mean peak value. These results held both for the adults and the children in our sample.[2] Note that the endpoint as we defined it is not always the minimum of the sentence; however, we found that it seemed to correspond in general to the lowest value of the speaker's range, as observed within our sample (See Table 1).

Discussion: Our research initially focused on the problem of interpreting clause-peak variation in these conversations (see Menn & Boyce 1978). However, Boyce noticed early on how stable the falling-contour endpoints were, and we began to wonder what the theoretical significance of the difference between these two parameters might be. We hypothesized that they were carrying different kinds of information.

Now, it has always been agreed that intonation carries many different kinds of messages. Traditionally, these have included attitudinal or affect messages, such as anger, affection, and fear (Bolinger 1978, Stevens 1972), speaker identity, and linguistic messages such as boundary and stress position (Fry 1958, Lehiste 1970).

Peak pitches have been assumed to vary according to the type of contour used, the degree of accent or emphasis placed on certain words, and the emotional overtones conveyed (Bolinger 1978). Endpoint pitch contributes to the gross differentiation among rising, steady, and falling contours; Bolinger, at least, seems to claim that falling-contour endpoint does vary, being lowered "to convey an attitude of finality (as in argument...)" or at the end of a "paragraph or discourse" (p.475, op.cit.).[2]

Our data cannot prove Bolinger wrong, because it is quite possible that our discourses do not contain the types of interchanges that he was referring to. But they do make it clear that we should construct an intonation theory in which fixed falling-contour endpoints are the norm, leaving large deviations from that norm as being highly marked cases.

So let us consider what such a theory would be like. What might be the communicative value of speaker-specific endpoints? What can we say about the different communicative roles of peaks and endpoints? In Menn & Boyce 1978 we presented evidence that peak variation from one clause to the next, regardless of the placement of the peak within the clause, is used to signal certain pragmatic and linguistic aspects of discourse structure, especially change of topic, pursuit of same objective, and back-channel ('contentless') responses vs. responses with message content. Although this was an interesting result, it brought up an even more interesting question. Why should clause-peak height, rather than some other variable such as clause-average F_o, be a clear indicator of discourse structure? Together with our data on endpoints, this finding seemed to favor a theory of intonation in which peak height is chosen for reasons external to the structure of the sentence, and in which endpoint values are chosen on the basis of the intonation contour type of the sentence.

At this point, we need to stop and acknowledge some problems inherent in the notions we are working with. First, we are still characterizing the significance of the measurements we are making in the subjective terms in which this area abounds: 'affect', 'style', 'excitement', 'disagreement', etc. However, we see this as a temporary intermediate stage: we have tried to formulate these hypotheses in such a way that our personal judgements can be be tested by perceptual experiments when vocoded or synthesized materials become available.

Second, linguistic/phonetic/musical literature is full of terms with overlapping meaning for even the simplest concepts about intonation. A brief discussion of terms may be in order. We make a distinction between the word contour, used to refer to generalized intonational shapes associated with certain broad categories of sentences, as in "declarative intonation contour", "wh-question contour", etc., and the word tune, used to refer to a member of a set (lexicon) of possible configurations which conveys a specific and specialized intonational meaning, such as 'contradiction'. Although we more or less adopt Liberman's notion that a lexicon of such 'tunes' exists (Liberman & Sag 1974, Ladd 1978a),we will be more concerned with the characteristics of contours. We use the word contour for the sake of convenience; this does not mean that we have taken sides in the 'levels' vs. 'contours' controversy (Bolinger 1951, Ladd 1978).

Consider a theory in which peak height reflects discourse structure and attitude, while endpoint values are chosen on the basis of the intonation contour type of the sentence except for completeness/incompleteness signals. Such a theory would account for all our findings. Summarized below, these are:

1) For rising contour sentence types, such as yes/no questions, the endpoint is often the maximum. The height of its rise is variable.

2) For falling-contour 'finished' sentence types, such as wh-questions and declaratives, endpoints are fixed and maximum peaks are variable.

3) Sentences with continuation rises (Maeda 1976, O'Shaughnessy 1976), sentences which are members of an itemized list, and sentences which for other reasons sound 'unfinished' stop above the preferred endpoint. (Note here that some authors have suggested that the major division among intonation contour types is that of 'finished' vs. 'unfinished' and that questions as well belong to the latter category (Lea, forthcoming)).

In addition, such a theory fits in well with what is known about the range of the normal human speaking-voice register, namely, that range is indefinitely more expandable in the higher regions than in the lower (Abramson, pers. comm.).

Our results are also compatible with recent work by J. Pierrehumbert (formerly Breckenridge) on synthesis of intonation. Without going too deeply into the details of this work, let us just sketch the salient points here.

If a line is drawn connecting the F_o peaks of an utterance and another line connecting the valleys, the area between is called the 'envelope' of the utterance. Although for individual sentences the slopes of both lines may vary (Pierrehumbert, pers. comm., M. Liberman, lecture) superimposing contours of the same type and duration (Henke, pers.comm.) reveals a typical envelope pattern. In a typical declarative envelope the top line drifts down at a much faster rate than does the baseline. In a typical yes/no question envelope the top line drifts up and the bottom line drifts down. Since yes/no question envelope patterns are practically unstudied to date, the following discussion will refer to declarative contours only.

Pierrehumbert's synthesis program is built on the envelope model. Keeping all other parameters, such as segmentals, duration, etc., of her 'model' speaker constant, she uses a slightly concave parabolic formula to derive topline envelope shape from preset peak parameters. These peaks are assumed to represent the highest stress levels in the sentence. Assignment of all other, secondary, stress accents is determined (in modified SPE tree fashion) relative to the highest pitch peaks. In order to do this she divides the envelope into horizontal regions, each corresponding to a different stress level. A diagram of this idealized envelope is presented in Fig. 2 . Note that, as the top-line drifts downward, each region narrows along with it.

This model is important for our theory for two reasons. First, envelope shape is determined by preset <u>peak</u> parameters. Of these, since the direction and relative amount of declination in a normal non-contrastive declarative sentence is quite predictable (Sorenson & Cooper 1978), the most important variable is the pitch of the highest peak. Second, Pierrehumbert found that, when the height of the preset F_o peak was varied, and the relative values of secondary peaks kept constant, listeners had difficulty hearing much difference between the variants, beyond a certain greater excitment or 'emphasis' that seemed to go along with higher peak height (Pierrehumbert, informal presentation in seminar).

Such a model, if accurate, suggests that the highest peak of a sentence may be the free-est variable a speaker has in his production of a particular sentence, assuming that he has already chosen the 'tune' and accentual pattern he wants to use. Again, this dovetails with our findings on the variability of clause-maximum F_o height (compared to endpoints) and our previously reported results on the correlation of clause-maximum F_o variation and discourse structure. It makes sense that a variable which is relatively free (i.e., unassigned) at a lower level of structure should be obligatorily assigned at a higher level.

It is not our intention, at the present time, to claim that this or any other model is the correct one for description of intonation behavior. We simply want to indicate that, of all the models we have seen to date, this one comes closest to capturing the generalizations implied by our data.

The present results bring up some interesting implications for the perception of intonation. Linguists have always had the following problem: if different gradients of pitch have such specific and subtle meanings for speakers, and different speakers have different vocal ranges, how does the hearer know how to interpret a certain pitch as meaning such-and-such in the speaker's system? We suggest that the hearer quickly determines the limits of the speaker's range, at least roughly, and then interprets contours with respect to those limits. Our results seem to suggest that the bottom line of this range is more-or-less fixed at the speaker's teminal fall endpoint, and that pitch interpretation may take place only in the higher limits of the range, as 'measurable (perceptible) degree of excursion from a fixed reference point'. This hypothesis is attractive in that it seems to simplify the process of intonation perception, since, in general, we do form expectations of usage norms for other aspects of language. It also has the advantage of being testable.

This model also provides a framework to account for one of the most prominent phenomena of speech addressed to children. It has been almost universally remarked that adults 'sound high' when they speak to very young children. It has also been remarked that the intonation contours used by adults are exaggerated, presumably for language-teaching purposes. Phillips (1970) showed that when the speech of mothers to children and mothers to other female adults was compared, mean minimum F_o remained the same, regardless of addressee, but that mean peak F_o was higher in the speech addressed to children. This result, together with our findings, indicates that when parents' voices 'sound high' in addressing children, it is because they raise their peak F_o. Now, suppose that there is a physiological 'floor' that essentially determines each person's preferred endpoint. Then, as A.W.F. Huggins has pointed out to us (pers. comm.), this would mean that if parents wish to exaggerate their contours in order to communicate more clearly, they must do so by raising their pitch peaks.

We can derive another theoretical result from our data. Musical notation has been advocated by a number of people as being appropriate for the description of intonation contour (a recent example is Gardiner 1978, summer LSA). There is no question that one can use pitch notation at least as well as sketches of curves or arrays of numbers to describe a particular contour. But, as we all know, there is more in the choice of notation than its ability to describe an event. If a notation is appropriate for a theory, it must make it easy to preserve the right invariances --- that is, if the theory says that two terms belong to the same class, the notation must make these two items look more alike than two items that the theory says belong to different classes. If a notation cannot capture this basic 'mapping' requirement, it is useless.

Musical notation recognizes two invariances: identity of pitch and identity of intervals. To our knowledge, it has never been suggested that pitch invariance is useful for the classification of contours. We would argue from our present findings that identity of intervals is not a useful invariant for the study of contours either --- at least, not according to the notion of 'sameness of contours' we described earlier.

Consider: Suppose we have two contours, one reaching a peak twice as high as the other (i.e., an octave higher). If each of these contours were composed of the same sequence of intervals, the endpoint of the first contour would have to be twice as high as the endpoint of the second contour. But we have shown that complete falling-contour sentences for the same speaker will have virtually the same endpoints regardless of the height of the peaks.

Therefore, either every peak height defines a different contour (contrary to any theory of contour that we know, and contrary to Pierrehumbert's findings with synthesis) or an interval-preserving theory is wrong.

Summary: We have one simple result: peaks vary a great deal more than falling-contour endpoints. In particular, we have confirmed sentence-by-sentence the fact that Phillips (1970) reported for means: the impression of high pitch in parents' speech to young children is a function of the higher pitch of sentence peaks. We have offered one firm interpretation that can be drawn from these results, namely, that the use of musical notation as a system for representing intonation contours can be ruled out. In addition, we have offered another, less firm, interpretation of the difference in variability between peaks and endpoints --- that is, that peaks and endpoints are carrying different types of information. We also proposed that parents' use of high pitch in speaking to children can be explained as a device to exaggerate their pitch signalling in the light of the constancy of endpoints.

The rest of what we offer is suggestions, hypotheses, or questions. Let us conclude by reviewing what we consider to be the major issues raised here, and the most interesting questions to explore in the future.

The major issues are:

1) Can the influence of discourse structure and speaker attitude on intonation be separated from such sentence-level phenomena as placement of stress and 'tune' assignment? And can the influences of discourse structure and speaker attitude be separated from one another?

2) What is the underlying explanation for the observed stability of falling-contour endpoints? Is it determined by a physiological floor? How consistent is it for each speaker on different days and in different situations? Whether it is stable from one session to the next or just stable for single sessions, to what degree is this stability physiologically automatic and to what degree is it a form of learned behavior? And if it is learned or is unstable from one session to the next, what social, cultural, or linguistic factors control it?

Questions of lesser scope include the following:

1) What are the factors (semantic, pragmatic, segmental, and intonational) that are sufficient to make an utterance sound complete? Which are necessary?

2) How much are listeners aware of the preferred endpoint of the speaker? How necessary is such awareness for interpretation of

the speaker's intonational meaning? What information is needed for the listener to form judgements of the expected range of the speaker? What would happen if an experimenter artificially manipulated sentence endpoints?

We have been assuming that our data on adult-child discourse is generalizable to adult-adult conversation. This seems probable in the light of Phillips' (1970) result that mean minimum F_o for mothers' (spontaneous) speech to children is the same as the mean for their (spontaneous) speech to another adult female, and the common-knowledge facts about adult 'read' speech. However, it's as well to remember that there may be systematic differences in endpoint/peak behavior between adult-adult and parent-child discourse that we don't know about. An analysis of such behavior in spontaneous adult interaction is obviously the next step.[3]

Finally, we would like to try to clarify some of the differences between our approach to 'intonational meaning' and that of the majority of linguists who have worked on it. The traditional linguist's approach is basically holistic, attempting to account for all the various uses of intonation in one, rather mystical, gestalt. A typical example is Pike, who saw intonation as "merely a shade of meaning added to or superimposed upon [the] intrinsic lexical meaning [of the sentence] " (1945, p.21); another is Liberman, who attempts to identify a 'lexicon' of intonational tunes whose abstract meaning is "intuitive" but whose specific meaning is defined only from the pragmatic circumstances of its use --- rather as one would describe the 'meaning' of a given word order (Liberman 1975, Ladd 1978a). Segmentation, in these cases, is used horizontally to separate out those sequences of tones, or contours, which seem to convey this vague, 'intuitive' meaning. We feel, however, that our results imply a different type of segmentation --- one that is more akin to the generative concept of 'levels of structure'. Of the two parameters that we have been working with, we have suggested that peak height would be principally determined at the level of discourse structure (Menn & Boyce 1978), while endpoint value would be determined primarily at the level where the contour is chosen (whether declarative, wh-question, yes/no question, etc.). In other words, we suggest some division into segment-level intonational rules, sentence-level rules, and discourse-level rules, something like stress-assignment rules in the SPE tradition.

* The research reported inthis paper was supported by NSF Grant no. BNS77-26871.

[1] There were four children in all, aged 2.1, 2.5, 2.6, and 4.0.

[2] One child's endpoints show unexplained scatter and will require further study (See Table 1).

[3] Bolinger's statement is somewhat ambiguous, in that he may be speaking of variation in 'depth of terminal fall' as due to vari-

ation in the slope of fall. Our results are relevant to either case.

[4] For instance, one difficulty is to account for the differences in standard deviation between Maeda's endpoint data and ours. Maeda had 3 male speakers read, in turn, a corpus of 30 sentences. Of these, the lowest standard deviation was 1.7 Hz. and the highest 2.8 Hz. Our lowest standard deviation for a male speaker was 8.2 and our highest 10.4. At this point, we have no way of knowing how much this difference is due to the difference between read and spontaneous speech, to normal intra-speaker variation over time (Atkinson 1973), to differences between parent-child and adult-adult interaction norms, or our own measurement error.

References:

Atkinson, J. (1973). Aspects of intonation in speech: implications from an experimental study of fundamental frequency. Unpublished U. of Connecticut doctoral dissertation.

Bolinger, D. (1951). Intonation: levels *vs.* configurations. *Word* 7: 199-210.

Bolinger, D. (1978). Intonation across languages. In: *Universals of human language, vol. 2 - Phonology*. J. Greenberg, ed. Stanford Univ. Press, Stanford.

Breckenridge [Pierrehumbert], J. (1977). Declination as a phonological process. Bell Laboratories Technical Memorandum, Dept. 1225.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech* 1:120-152.

Gardiner, D. B. (1978). Intonation in musical notation. Paper read at the summer meeting of the Linguistic Society of America, July 1978.

Garnica, O. K. (1977) Some prosodic characteristics of speech to young children. Ohio State University working papers in linguistics, Ohio State University. (Revised version of 1975 Stanford University doctoral dissertation.)

Ladd, D. R. (1978a). The structure of intonational meaning. Unpublished Cornell University doctoral dissertation.

Ladd, D. R. (1978b). Stylized intonation. *Lg 54*(3), 517-540.

Lea, W. (forthcoming). *Prosodic aids to speech recognition*. Englewood Cliffs, N.J.: Prentice-Hall.

Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge, Mass: MIT Press.

Liberman, M., and I. Sag (1974). Prosodic form and discourse function. *In* Papers from the 10th regional meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society. 416-427.

Liberman, M. (1975). The intonational system of English. Unpublished MIT doctoral dissertation. 30-32.

Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. Linguistic Inquiry 8:249-336.

Maeda, S. (1976). A characterization of American English intonation. Unpublished MIT doctoral dissertation.

Menn, L., and Boyce, S. (1978). Time course of fundamental frequency variation in parent-child discourse. Paper read at the annual meeting of the Linguistic Society of America, 1978.

O'Shaughnessy, D. (1976). Modelling fundamental frequency, and its relatio to syntax, semantics, and phonetics. Unpublished MIT doctoral dissertation.

Pike, K.L. (1945). The intonation of American English. Ann Arbor: U. of Michigan Press.

Pierrehumbert, J. B. (1978), ms. The perception of intonation. (To appear in J. Acoustic Soc. of America.)

__________________ (forthcoming). Intonation synthesis based on metrical grids. (Available in preprint volume of the Acoustical Society of America meeting of June, 1979.)

Phillips, J. R. (1971). Formal characteristics of speech which mothers address to their young children. Unpublished doctoral dissertati Johns Hopkins University. Distributed by Xerox University Microf

Sorenson, J. M., and W. E. Cooper (1978). Syntactic coding of fundamental frequency in speech production. Chapter to appear in R.A. Cole (Perception and production of fluent speech. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Williams, C.E., and K.N. Stevens (1972). Emotions and speech: some acoustic correlates. J. Acoustic Soc. of America 52: 1238-1250.

FIG.1. SCHEMATIZED DRAWING OF TYPICAL INTONATION CONTOURS SHOWING DEFINITION OF "F_0 PEAK" AND "ENDPOINT OF FALLING-CONTOUR SENTENCE"

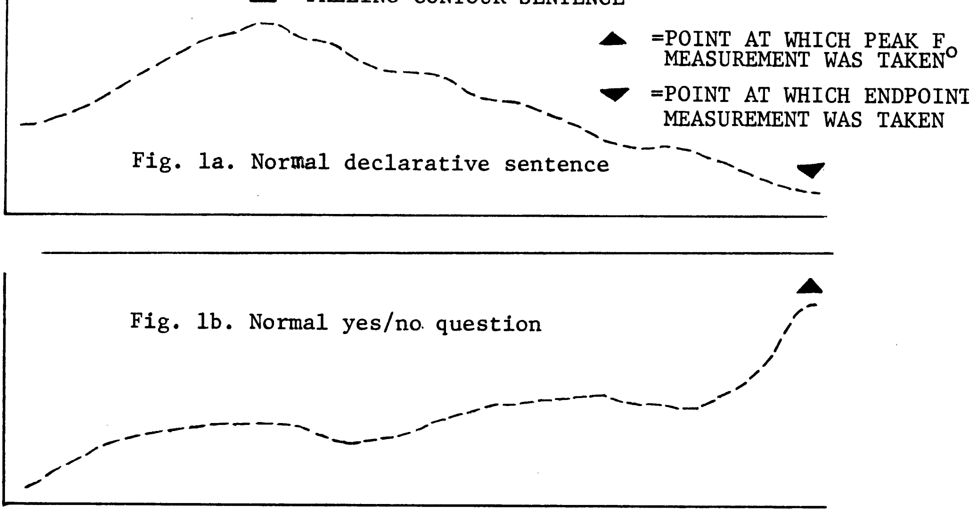
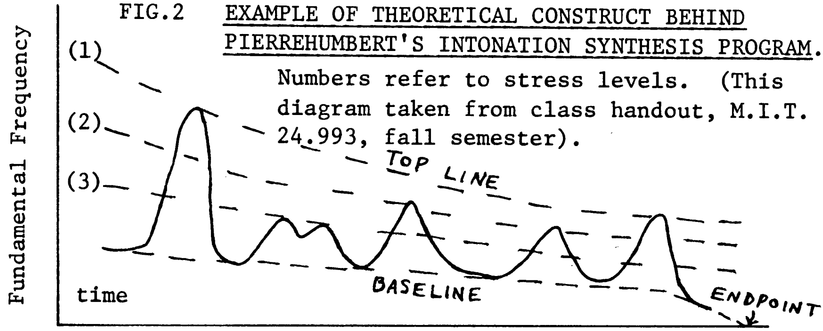


FIG.2 EXAMPLE OF THEORETICAL CONSTRUCT BEHIND PIERREHUMBERT'S INTONATION SYNTHESIS PROGRAM. Numbers refer to stress levels. (This diagram taken from class handout, M.I.T. 24.993, fall semester).



In (1) November, (3) the (3) region's (2) weather was (2) unusually (1) dry.

DISCOURSE		PEAKS					ENDPOINTS				
		#	MEAN	σ	GEO.MEAN	GEO. σ	#	MEAN	σ	GEO.MEAN	GEO. σ
Child #14 (2 years 1 month)	MOTHER	70	320.8	73.8	313.1	1.242	33	188.8	12.9	188.3	1.071
+ Mother	CHILD	21	460.2	169.1	441.9	1.290	13	251.2	22.5	250.2	1.094
Child #14	FATHER	88	195.2	64.5	186.1	1.352	30	102.8	10.4	102.4	1.102
+ Father	CHILD	36	392.4	77.3	384.4	1.230	21	242.8	11.8	242.5	1.050
Child #16 (2 years 6 months)	MOTHER	50	344.8	109.4	328.6	1.359	22	203.3	15.3	202.8	1.078
+ Mother	CHILD	24	358.8	95.9	346.8	1.296	11	229.5	14.7	229.0	1.070
Child #16	FATHER	62	148.6	37.6	144.0	1.329	37	91.5	10.4	90.9	1.125
+ Father	CHILD	44	425.1	78.7	417.3	1.224	28	259.8	7.8	259.7	1.030
Child # 22 (4 years)	MOTHER	65	300.3	83.4	286.7	1.392	31	200.2	27.1	198.6	1.126
+ Mother	CHILD	20	369.3	83.1	361.6	1.218	17	257.6	35.8	255.4	1.139
Child # 13 (2 years 5 months)	FATHER	58	203.9	46.4	198.7	1.255	14	102.2	8.2	101.9	1.084
	CHILD	27	327.6	52.1	323.1	1.185	22	206.5	49.8	181.1	2.200

TABLE 1. Collected data on 6 speaker pairs, including number of tokens, arithmetic and geometric (logarithmic) means, and arithmetic and geometric (logarithmic) standard deviations for both peak and endpoint F_0 measurements.